

## Portland State University PDXScholar

---

Dissertations and Theses

Dissertations and Theses

---

1-1-2010

# Diatom-Based Stream Bioassessment: the Roles of Rare Taxa and Live/Dead Ratio

Nadezhda Dimitrova Gillett  
*Portland State University*

Let us know how access to this document benefits you.

Follow this and additional works at: [http://pdxscholar.library.pdx.edu/open\\_access\\_etds](http://pdxscholar.library.pdx.edu/open_access_etds)

---

### Recommended Citation

Gillett, Nadezhda Dimitrova, "Diatom-Based Stream Bioassessment: the Roles of Rare Taxa and Live/Dead Ratio" (2010).  
*Dissertations and Theses*. Paper 26.

[10.15760/etd.26](https://doi.org/10.15760/etd.26)

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

Diatom-Based Stream Bioassessment: the Roles of Rare Taxa and Live/Dead Ratio

by

Nadezhda Dimitrova Gillett

A dissertation submitted in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy  
in  
Environmental Sciences and Resources

Dissertation Committee:  
Yangdong Pan, Chair  
Joseph Maser  
Alan Yeakley  
Ian Waite  
Jiunn-Der Duh

Portland State University  
©2010

## Abstract

It is unclear whether accounting for the number of rare taxa or differentiating live (cells with visible chloroplasts) and dead (empty cells) diatoms would enhance the accuracy and precision of diatom-based stream bioassessment. My dissertation research examines whether the number of rare taxa and percentage (%) live diatoms can be used as indicators of human disturbance in streams/rivers. To address my objectives, I analyzed two datasets collected at different spatial scales. The large scale dataset was collected over the course of five years and included more than 1300 sites and over 200 environmental variables from 12 Western US states (US EPA's Environmental Monitoring and Assessment Program). The small scale dataset included 25 streams from the Oregon Coast Range. My findings revealed that rarity definition is important in bioassessment. Rural taxa richness (number of taxa with high occurrence and low abundance) was the only rarity metric that distinguished least disturbed (reference) sites from the most disturbed (impacted) ones. Neither of the other two rarity metrics, satellite (taxa with low occurrence and abundance) and urban (taxa with low occurrence and high abundance) taxa richness, was able to do that. Results from regression tree analysis revealed that rural taxa richness increased with human disturbance, but it was significantly higher at impacted sites only in the Mountains ecoregion (t-test,  $p < 0.05$ ). Neither of the other two ecoregions (Plains and Xeric) exhibited significant differences between reference and impacted sites (t-test,  $p > 0.05$ ). Percentage live diatoms distinguished reference from impacted sites only in

the Mountains ecoregion (t-test,  $p=0.02$ ) and somewhat in the Plains (t-test,  $p=0.05$ ). However, % live diatoms exhibited opposite patterns in the two ecoregions. They increased with disturbance in the Mountains and decreased in the Plains. The results from the small scale study in the Oregon Coast Range revealed similar species compositions between live and live+dead diatom assemblages (non-metric multidimensional scaling) and similar relationships with their environmental variables (linear fitting). Both assemblages correlated well with in-stream physical habitat conditions (e.g., channel dimensions, substrate types, and canopy cover). Both rural taxa richness and % live diatoms can be used as indicators of human disturbance in streams/rivers, especially in the Mountains ecoregion, dependent on specific assessment objectives.

To my family

## **Acknowledgements**

I greatly appreciate the help, understanding and patience of all people involved with this work. I am grateful to all of my committee members Dr. Joe Maser, Dr. Alan Yeakley, Dr. Ian Waite, and Dr. Jiunn-Der Duh for their expertise, guidance and support. I would like to express my deepest gratitude to my dissertation advisor, Dr. Yangdong Pan, for all of his guidance, patience and understanding. I am greatly indebted to him for all of my knowledge about aquatic ecology, bioassessment and data analysis. Thanks for all scientific, psychological, personal and other advice.

Thanks to many colleagues from Pan's lab for their assistance with field and lab work. Namely, Christine Weihoefer, Xiaoyan Wang, Miguel Estrada, Chris Parker, Patrick Edwards, Jeff Meacham, and Hsiao-Hsuan Lin. Special appreciation is due to Rosalina Hristova and Kalina Manoylov for all professional and emotional support. Thanks to all ESM faculty, staff and students for help and assistance. I am grateful for the opportunity to teach in the Department and receive travel awards that allowed me to attend scientific meetings. Thanks to the Office of Graduate Studies for one-term tuition waver.

Many thanks to my family, especially to Mom and Dad, for waiting patiently to finally see me a Doctor. Thanks to my sister, Maria, for all possible forms of support. Special thanks to my husband, Chris, who has been my rock all the way through this adventure, who took great care of our family, and who never once complained about anything. Finally, thanks to my son, Mitko, who has been the Sunshine in my life and who was making me smile even when I did not feel like it.

## **Table of Contents**

Abstract.....	i
Dedication.....	iii
Acknowledgements.....	iv
List of Tables.....	vi
List of Figures.....	viii
Introduction.....	1
Background Information.....	4
Chapter 1: The importance of rare taxa in diatom-based stream bioassessment.....	14
Chapter 2: The importance of live/dead diatoms in bioassessment: a large scale study of Western US streams.....	53
Chapter 3: The importance of live diatoms in bioassessment: a regional study of Western US streams.....	91
Summary, Implications and Recommendations for Water Quality Assessment.....	115
References.....	119
Appendix .....	136

## List of Tables

Table 1.1. Regional rarity classes.....	22
Table 1.2. Sample characteristics and number of taxa in each rarity class with mean (minimum and maximum) values for the whole dataset (West), three ecoregions (MT-Mountains, PL-Plains, XE-Xeric), reference and impacted sites.....	30
Table 1.3. Number of taxa and mean relative abundance (minimum-maximum) for regional rarity classes.....	32
Table 1.4. Summary of the regression tree (RT) models used to associate variation in rarity metrics with environmental features (predictors) for the whole dataset (West), two ecoregions (MT-Mountains and PL-Plains) and reference sites. No models were developed for Xeric ecoregion and impacted sites. Number of nodes and variables are measures of the complexity of the models. $r^2$ measures the strength of association between the metric and the predictors. Metric descriptions and abbreviations as in text and Table 1.1. Variable abbreviation and summaries are in the Appendix.....	34
Table 1.5. Weighted average values, optima (tolerances), number of sites and mean relative abundances for selected rare and common taxa. Variable abbreviation and summaries are in the Appendix .....	43
Table 2.1. Environmental variables description and summaries for the whole dataset (West), three ecoregions (MT-Mountains, PL-Plains, XE-Xeric), and two stream order classes (1st-3rd and > 3rd).....	66



Table 2.2. Summary for sampled sites (n) and % live diatoms in the dataset (West), three ecoregions (MT-Mountains, PL-Plains, XE-Xeric) and two stream order classes (1st-3rd and > 3rd).....	67
Table 2.3. Results from multiple regression for the whole dataset (West), two ecoregions (MT-mountains and XE-xeric) and two stream order classes (1st- 3rd and > 3rd). No model was developed for the PL (Plains). “+” and “-“ signs represent significant positive or negative relationship, respectively. Numbers in parentheses denote number of sites.....	78
Table 3.1. Minimum, mean and maximum values for some environmental variables for Oregon Coast Range streams (n=25).....	101
Table 3.2. Diatom community characteristics with mean (minimum and maximum) values.....	102
Table 3.3. Squared correlation coefficient ( $r^2$ ) and its significance between the selected environmental variables and the three assemblages (LD, L and D).....	108

## List of Figures

Figure 1. Map of WEMAP study area and three spatial scales: West-wide (12 states), three climatic zones (Mountains, Plains, Xeric) and ten aggregated Level III ecoregions (Omernik, 1987) (from <a href="http://www.epa.gov/emap/west">http://www.epa.gov/emap/west</a> ).....	10
Figure 2. An example of EMAP site selection process in the Mid-Atlantic region. A map of stream network overlaid with 40 km <sup>2</sup> hexagons (from Herlihy et al., 2000). See text for explanation.....	11
Figure 3. WEMAP sampling reach features. The sampling site (X-site) lays in the center of the sampling reach which is defined by 11 equally spaced transects (A-K). Biological data are collected on the left (L), center (C) or the right (R) side of the stream, after a random start. Transects delineate stream reach which equals the length of 40 times the mean wetted width at the X-site. See text for details.....	12
Figure 1.1. Maps of USA and 12 Western states with locations of sampling sites (n=987).....	21
Figure 1.2. Regression tree building: a) an example of the terminology of a regression tree; b) “growing” a regression tree, $D_i$ – deviance in the response variable at the root of the tree, $D_{i,L}$ - deviance in the response variable in the left new group, $D_{i,R}$ - deviance in the response variable in the right new group, $i$ – response variable.....	24

Figure 1.3. Regression tree cross-validation with multiple trees of different sizes (a)	
and selection of final regression tree with the lowest mean prediction error	
(b).....	25
Figure 1.4. Plot of species occurrence and mean % relative abundance (RA) in the	
Western US streams/rivers. Each point represents a diatom taxon. Dashed lines	
define boundaries between rarity classes with number of taxa in them (numbers	
in parentheses).....	32
Figure 1.5. Final regression trees for rural (a, $r^2=0.32$ ) and core (b, $r^2=0.36$ ) metrics.	
Names and numbers above each node indicate the variable and its value used	
for the split. Sites which meet the node criteria are split to the left. Names	
below the first split indicate alternative variables. Numbers at the leaves	
specify mean predicted values for the response variable (rural and core	
metrics) and the number of sites (n). Variable descriptions are in the	
Appendix.....	35
Figure 1.6. Site summaries for the left and the right nodes on Fig. 1.5a.....	36
Figure 1.7. Rural metric at reference and impacted sites in the dataset and the three	
ecoregions (MT, Mountains; PL, Plains; XE, Xeric) (a) and in each of the	
ecoregions divided by stream size and diatom successional stage: MT (b), PL	
(c) and XE (d). Numbers in parenthesis show the number of sites.....	38
Figure 1.8. Boxplots of coefficient of variation (weighted average tolerance/ weighted	
average optima) for five core (C), rural (R), urban (U) and satellite (S) taxa in	
the whole dataset (a) and in the MT ecoregion (b). Their names and summaries	

are provided in Table 1.5. Metric descriptions are in Table 1.1. Variable abbreviations and summaries are in the Appendix.....	42
Figure 2.1. Conceptual model for the influence of different environmental factors on the abundance of live diatoms (modified after Stevenson, 1997).....	57
Figure 2.2. Maps of USA and sampling sites (n=587) in 12 states in the Western USA.....	59
Figure 2.3. Non-metric multidimensional scaling plot for diatom species composition between two within-year site visits (n=27). The arrow length corresponds to Bray-Curtis similarity distance. Arrows point in the direction of the second site visit.....	68
Figure 2.4. Boxplot for the difference in % live diatoms between 2 <sup>nd</sup> and 1 <sup>st</sup> within-year site visits (n=27).....	69
Figure 2.5. Boxplots for early and late diatom successional sites (numbers in parentheses) in the dataset (West) and three ecoregions (MT-Mountains, PL-Plains and XE-Xeric) with p-values from t-tests.....	70
Figure 2.6. Boxplots for selected environmental variables from the conceptual model for early and late diatom successional sites (numbers in parentheses) in the MT ecoregion.....	71
Figure 2.7. Boxplots for early and late diatom successional sites (numbers in parentheses) at low (1 <sup>st</sup> -3 <sup>rd</sup> ) and high (>3 <sup>rd</sup> ) order streams in three ecoregions (MT-Mountains, PL-Plains and XE-Xeric) with p-values from t-tests.....	72

Figure 2.8. Boxplots for selected environmental variables from the conceptual model for early and late diatom successional sites (numbers in parentheses) in small (1 <sup>st</sup> -3 <sup>rd</sup> order) streams in the MT ecoregion.....	73
Figure 2.9. Boxplots for invertivore fish and macroinvertebrate grazers density at sites (numbers in parentheses) with and without piscivore fish with p-values from t-tests.....	75
Figure 2.10. Boxplots for % live diatoms and chlorophyll <i>a</i> concentrations at sites (numbers in parentheses) with and without piscivore fish with p-values from t-tests.....	76
Figure 2.11. Boxplots for reference and impacted sites (numbers in parentheses) in three ecoregions (MT-Mountains, PL-Plains and XE-Xeric) with p-values from t-tests.....	80
Figure 2.12. Boxplots for selected environmental variables from the conceptual model for reference and impacted sites (numbers in parentheses) in the MT ecoregion.....	81
Figure 2.13. Boxplots for reference and impacted sites (numbers in parentheses) at low (1 <sup>st</sup> -3 <sup>rd</sup> ) and high (>3 <sup>rd</sup> ) order streams in three ecoregions (MT-Mountains, PL-Plains and XE-Xeric) with p-values from t-tests.....	82
Figure 2.14. Boxplots for selected environmental variables from the conceptual model for reference and impacted sites (numbers in parentheses) in small (1 <sup>st</sup> -3 <sup>rd</sup> order) streams in the MT ecoregion.....	83

Figure 3.1. Map and location of the sampling sites in the Northern Oregon Coast Range, OR, USA.....	96
Figure 3.2. Scatterplots for different diatom community characteristics: (A) species richness of LD and L, (B) species richness of D and L, (C) species richness of LD and percent similarity between LD and L, (D) percent live diatoms and percent similarity between LD and L. The lines in A and B are 1:1 ratio. * significance at $p < 0.05$ .....	103
Figure 3.3. Scatterplots for the percent relative abundances of the dominant diatom species in LD and L assemblages with a 1:1 ratio line. Refer to Table 3.2 for complete species names. * significance at $p < 0.05$ .....	104
Figure 3.4. Scatterplots for the percent relative abundances of the dominant diatom species in D and L assemblages with a 1:1 ratio line. Refer to Table 3.2 for complete species names. * significance at $p < 0.05$ .....	105
Figure 3.5. Bubble plot for the distribution of relative abundance of <i>Achnantheidium minutissimum</i> among the sites for the LD (A) and L (B) counts. Bigger bubble size corresponds to higher relative abundance.....	106
Figure 3.6. Plot of the residuals from procrustes analysis for the comparison between the ordinations of LD and L. Each arrow connects the LD and the L count from one site. The arrows point in the direction of the L ordination. The arrow length corresponds to the difference in site location between the two ordinations.....	107

Figure 3.7. Non-metric multidimensional scaling plots for the LD (A) and L (B)

assemblages, and significant environmental variables ( $p < 0.05$ ) from the linear fitting function. Abbreviations: CanC – canopy cover, Chl – chlorophyll *a*, DecC – percent riparian deciduous canopy, Fines – percent fine substrates (silt/clay/muck), Sand – percent sandy substrates, SB – percent substrates as smooth bedrock, Width – mean wetted width..... 109

## **Introduction**

The assessment of ecological integrity in streams and rivers is essential for the healthy, social and economic well-being of humans. However, human-induced stressors (e. g. agriculture, logging, urbanization, sewage contamination, etc.) alter the stream environment and the aquatic organisms by changing their habitat structure, flow regime, water quality, energy source and biotic interactions (Karr and Chu, 1999). Conceptually such alterations may cause the biota to respond in a predictable way by changing its energetics, nutrient cycling and community structure (Odum, 1985). Bioassessment focuses on changes in biotic community structure and function in response to human disturbance.

Diatom-based assessment has been an integral part of both national and regional bioassessment programs (Stevenson and Bahls, 1999; Peck et al., 2006). Diatoms, a group of algae, are good bioindicators of water quality in streams and rivers (Stevenson and Pan, 1999). As primary producers, they are at the base of the food web in streams. Diatoms, with short life cycles, respond quickly to a variety of environmental conditions such as organic pollution (Lange-Bertalot, 1979), siltation (Bahls, 1993), pH (Pan et al., 1996) and eutrophication (Kelly and Whitton, 1995; Potapova and Charles, 2007). Diatom assemblages are species-rich and each species may provide information about their environment (Lowe and Pan, 1996).

Conventional diatom-based bioassessment programs, however, often draw assessment conclusions based on the responses of the common taxa to the environment, while the rare taxa in the diatom assemblages, regarded as noise, are often excluded from final



assessment analyses or down-weighted (Pan et al. 1996, 2000; Potapova and Charles, 2002, 2003). Based on the niche theory, several researchers assume that rare species may have narrow environmental requirements and therefore may be used as good indicators of pollution (Gaston, 1994; Potapova and Charles, 2004), yet this notion is not well supported by the current diatom literature.

In general, bioassessment uses data collected from living organisms to evaluate environmental quality, with diatoms being an exception. A diatom sample comprises a mixture of live (chloroplast containing siliceous cell walls) and dead (only siliceous cell walls) individuals which are not distinguished from each other in the conventional diatom analysis (Charles et al., 2002). Laboratory analysis of diatoms involves the use of strong acid to remove the organic material in the cells and to reveal the fine morphological features of their cell walls, which allows identification to the lowest possible taxonomic level. It is unclear whether differentiating live and dead diatoms in an assemblage would significantly improve the accuracy and precision of diatom-based bioassessment. Several researchers have suggested that assemblages generated by the traditional method may better integrate spatial and temporal variability of environmental conditions in the sampled stream and its watershed, because the accumulation of live and dead diatoms over time may provide broad environmental information (Cox, 1998; Stevenson and Pan, 1999). However, Pryfogle and Lowe (1979) concluded that the traditional counting technique might mask the specific response of the live resident taxa to the local environmental conditions. In addition, dead diatoms from far away or adjacent habitats might be a significant source of error

in characterizing diatom communities and their signals about the local environment (Round, 1998).

Data from the US Environmental Protection Agency's Western Environmental Monitoring and Assessment Program (WEMAP) provides a great opportunity to enhance the quality of diatom-based assessment. WEMAP sampled Western US streams and rivers with the major objective to assess the status of surface waters in the west. The project collected an extensive set of physical, chemical and biological data over the course of five years in 12 western states. Another important task of this project is to develop and evaluate the tools for assessing ecological conditions. Both of these objectives, as well as its large spatial coverage and probability-based sampling design, allow inferences to be made from a small number of samples about the extent of the whole study area. The extensive study area provides potential good natural and disturbance gradients which can be used to explore the roles of live/or dead diatoms and rare diatom taxa in bioassessment.

My research focuses on the potential roles of both rare taxa and live/dead diatoms in diatom-based stream bioassessment. Specifically, I want to evaluate whether the number of rare taxa can be used as an indicator of human disturbance. I also want to assess the relationships between live/dead diatoms and human disturbance from a large scale study (12 Western US states) and a small scale study (Oregon Coast Range).

## **Background Information**

### **Bioassessment**

The main goal of bioassessment is to evaluate whether a site or a system of interest (test site) is impaired based on biological indicators. This is usually done by a comparison to reference conditions with high physical, chemical and biological quality. The reference conditions can be defined for a region or for a site depending on the assessment objectives (US EPA, 2002). The regional reference condition represents the natural state of biological communities within a region for ecosystems with different geographical features (e.g., climate, soil, topography and vegetation) (Hughes et al., 1986). The site-specific approach evaluates the conditions upstream from a point source pollution or in a matching watershed (Hughes et al., 1986). Due to the pervasive nature of human disturbance, a reference condition free of human disturbance may not exist and the phrase “reference condition for biological integrity” (biological condition without human disturbance in a watershed, e.g., national parks) has been proposed (Stoddard et al., 2006). Several definitions for reference conditions include: minimally disturbed condition (absence of significant human disturbance), least disturbed condition (least impacted current condition), historical condition (known lack of human influence in the past) or best attainable condition (desirable future condition with better management) (Stoddard et al., 2006). Reference site selection is often determined by proximity and influence of human population density and distribution, road density and land use in the watershed (logging, mining, agriculture, grazing and urbanization). Biological indicators, organisms with well-

known response to disturbance, are often used to detect deviations from reference conditions (Davies and Jackson, 2006).

Bioassessment uses different groups of organisms (e.g., algae, macroinvertebrates and fish) to evaluate the environmental conditions of waters. Biological assemblages are often characterized using different biological attributes such as diversity, abundance, and functional groups. Diversity measures play an important role in bioassessment, because diversity is perceived as a correlate of environmental well being (Magurran, 2004). Diversity can decrease with disturbance, increase or have mixed response depending on the type of pollution (Patrick et al., 1954; Patrick, 1967; Stevenson, 1984, 2006). Under stressed conditions, species richness may decrease while dominance increases (increase in unevenness of species abundances) (Patrick, 1973; Odum, 1985). Many statistical models assume that the abundance of an organism is related to its ecological importance (Magurran, 2004). Functional groups are based on trophic level (e. g., herbivores, omnivores, carnivores, predators) (Karr, 1981), feeding habits (e. g., filterers, collectors, gatherers) (Vannote et al., 1980) or other processes that different community members perform. In addition, species composition of biological assemblages is used in a variety of bioassessment methods.

Major approaches in bioassessment

#### *Multimetric model*

The two major approaches in bioassessment are multimetric and multivariate. The multimetric approach uses multiple characteristics (metrics) of biological

communities to describe their community structure and function. Community structure includes species composition and richness of the whole community or individual groups (e. g., total number of species, number of sensitive or tolerant species) (Karr, 1981). Functional metrics incorporate the role that different members play in the community (e. g., trophic interactions) (Karr, 1981). Development of a multimetric index involves combining a number of individual metrics. Each metric is a hypothesis about the effects of human disturbance on the biological community and is assigned a score (1, 3 or 5) based on its response to disturbance (e. g., 1 will be assigned to a site with strong deviation from reference condition, while a 5 will be assigned to a site with little or no deviation from reference condition). For example, the number of fish species is expected to decrease with impairment (Karr, 1981), and a site with few species will receive a score of 1, while a species-rich site will receive a score of 5. The scores for all metrics for a site are summed up into an index of biotic integrity (IBI). This index is then used to assign the site to a biotic integrity class ranging from excellent (high IBI) to poor (low IBI) (Karr, 1981). The properties of an effective IBI include: selection of the right biological attributes (i. e., any measurable part or process of a biological system) which should be easy to measure and interpret (Karr and Chu, 1999); the attributes should increase or decrease predictably with increasing human disturbance and they should be sensitive to different ecological stresses. The IBIs should also be able to discriminate between human-caused and natural variations (Karr, 1999). IBIs have been developed for fish (Karr, 1981; Fore et al., 1994), macroinvertebrates (Kerans and Karr, 1994; Barbour et al., 1996; Fore et al., 1996)

and diatoms (Bahls, 1993; Hill et al., 2000; Fore, 2002; Wang et al., 2005). Some of the advantages of the multimetric approach include ease of understanding and application by non-specialists and managers by simple comparison of a single score to a target value (Reynoldson et al., 1997), and IBIs can incorporate biological and ecological information about an organism's habitat, feeding habits and reproduction. The critics of this approach point out its circularity (e. g., a site is assessed as impaired or not and then metrics are developed to evaluate it as degraded or not) and incomplete use of all data (e. g., not all biological attributes are used as metrics). Another criticism of IBIs includes redundancy caused by the use of multiple similar metrics and their tendency to compound error. A few additional drawbacks include arbitrariness in metric selection and scoring, difficulty in calibration and unknown sensitivity. Finally, the multimetric approach is criticized for its oversimplification of a complex system to one number which might result in loss of information (Suter, 1993; Reynoldson et al., 1997).

#### *Predictive model*

The multivariate approach compares the biota observed (O) at a site to the biota expected (E) to be there if the site is not impacted by human activity (O/E). This method predicts the biota expected at a site from a set of environmental variables which are unlikely to be influenced by human action (e. g., latitude and elevation) (Hawkins et al., 2000). The development of a predictive model includes three major steps. First, the species composition from all reference sites in an area of interest is used to assign the sites to groups using a classification technique (e. g., cluster

analysis). Second, the environmental variables measured at those sites are used to discriminate between the existing site clusters based on species composition. Third, the frequency of occurrence of a given taxon in a group is estimated (expected taxon, E). Then, when a new site (test site) is sampled, its environmental variables are matched with the ones from the predictive model and the site is assigned to a group. Knowing the frequency of occurrence of all the species in this group and the observed species in the test site (O), the model estimates the probability of capture of an observed taxon. O is the number of observed taxa at the test site which are also on the list of expected taxa (E). E is estimated as the sum of capture probabilities predicted by the model. The ratio of O over E (O/E) provides a measure of how many taxa from the test site are also observed in reference conditions and whether any of them might be lost to degradation.  $O/E \sim 1$  indicates reference conditions, while any departure from 1 signals impairment. However, O/E may not be 1 even when a test site is truly a reference site, because of the existing natural variation among the reference sites. The power of a predictive model is measured by its ability to assign correctly test sites to existing groups (Hawkins et al., 2000). The multivariate approach requires no assumptions about the grouping of reference sites and their comparison to test sites (Reynoldson et al., 1997), and is widely used by government agencies in the UK (River InVertebrates Prediction And Classification System, RIVPACS; Wright et al., 2000), Australia (AUStralian RIVer Assessment System, AUSRIVAS; Marchant et al., 1997) and Canada (BEnthic Assessment of SedimenT, BEAST; Reynoldson et al., 1995). Some drawbacks of the multivariate approach include complexity of the model

development (however, there are available software programs with easy-to-interpret results) (Reynoldson et al., 1997) and difficulty associated with the finding of undisturbed sites (Stoddard et al., 2006). This approach is also criticized for the omission of the biology of organisms (Karr, 1999). It has been suggested that the simultaneous use of multimetric and multivariate approaches might lead to better assessment decisions (Reynoldson et al., 1997).

#### Western Environmental Monitoring and Assessment Program (WEMAP)

##### *Sampling design and site selection*

The US Environmental Protection Agency's Western Environmental Monitoring and Assessment Program (WEMAP) sampled streams/rivers in 12 western states (Stoddard et al., 2005). The program collected biological, chemical and physical data at over 1300 stream and river locations during the low flow period (May through October) of 2000-2004. The results would provide information about the biological quality of flowing waters at different spatial scales: across the West (12 states), three climatic zones (Mountains, Plains, Xeric), and ten aggregated Level III ecological regions (Fig. 1) (Omernik, 1987; US EPA, 1996).



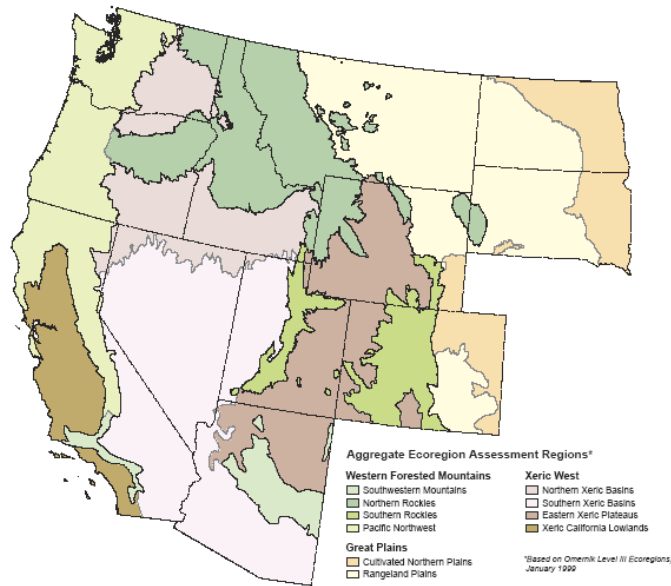


Figure 1. Map of WEMAP study area and three spatial scales: West-wide (12 states), three climatic zones (Mountains, Plains, Xeric) and ten aggregated Level III ecoregions (Omernik, 1987) (from <http://www.epa.gov/emap/west>).

A key element of the WEMAP is its probability-based sampling design for site selection. This feature allows for the results from a small number of samples to be applied to the whole area of interest with a known level of uncertainty. Sampling sites were selected from digital 1:100,000-scale USGS hydrographic maps of perennial and non-perennial streams and rivers following spatially-balanced stratified random sampling design (Herlihy et al., 2000; Hughes et al., 2000; Stevens and Olsen, 2004). A randomly placed grid of triangular points and their hexagonal tessellations were overlaid with the map of the continental US, allowing for a finer or coarser grid depending on the desired sample size and spatial coverage (Herlihy et al., 2000). Sampling sites were identified after a two-stage selection process. In the first stage, 40 km<sup>2</sup> hexagons were delimited and the length of streams within each one of them was

calculated (Fig. 2). In the second stage, individual sampling sites were selected from a subset of the first-stage sample after examination of maps, aerial and satellite imagery, and other types of data (Herlihy et al., 2000). All stream segments from the first-stage process (within a hexagon) were randomly arranged in a straight line and sampling points were randomly selected at equal intervals after a random start. The length of the intervals was determined by the length of the line divided by the number of desired samples. The total sampled stream length was 304,544 km or 48% of the original frame length, and 72% of the target stream length (Stoddard et al., 2005).

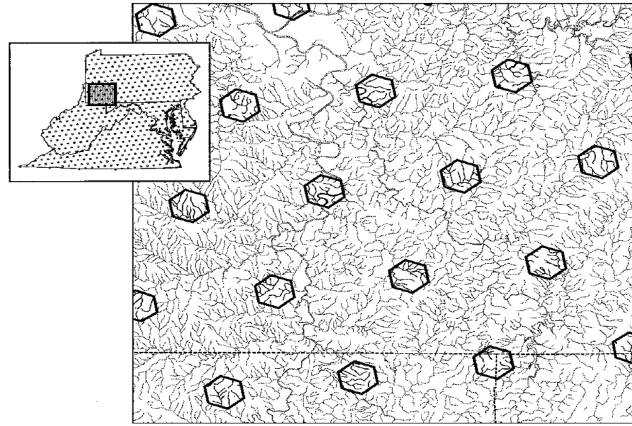


Figure 2. An example of EMAP site selection process in the Mid-Atlantic region. A map of stream network overlaid with 40 km<sup>2</sup> hexagons (from Herlihy et al., 2000). See text for explanation.

### *Field sampling*

Streams were sampled with wadeable-stream protocols (Peck et al. 2006), whereas rivers were sampled with rafts and boatable-river protocols (Peck et al., in press). WEMAP employed standardized reach length which increased in proportion to stream size. Each sampling site (X-site on Fig. 3) was positioned in the center of the sampling reach whose length is 40x the average stream wetted width or a minimum of

150 m (Kaufmann et al., 1999). Eleven equally spaced transects (A-K, Fig. 3) were delineated along the sampling reach.

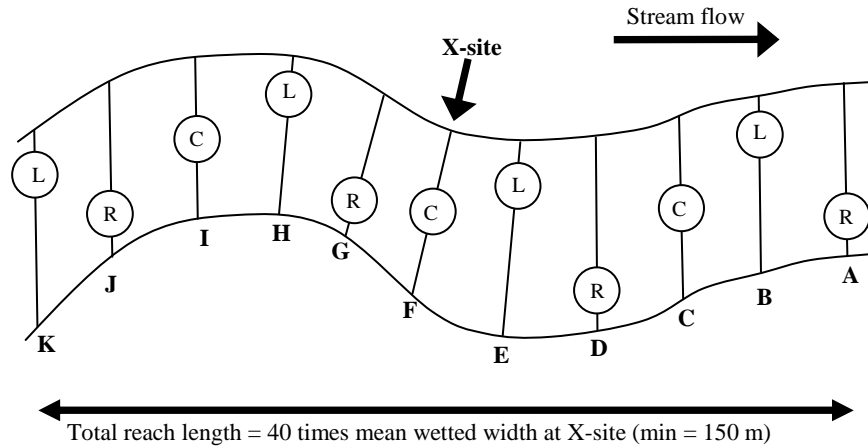


Figure 3. WEMAP sampling reach features. The sampling site (X-site) lays in the center of the sampling reach which is defined by 11 equally spaced transects (A-K). Biological data are collected on the left (L), center (C) or the right (R) side of the stream, after a random start. Transects delineate stream reach which equals the length of 40 times the mean wetted width at the X-site. See text for details.

A grab sample from the middle of the stream was collected for water chemistry analysis. It was used to determine acid-base status, trophic condition (nutrient enrichment), chemical stressors and classification of water chemistry type (Herlihy, 2006). Water chemistry measurements included major cations ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ) and anions ( $\text{Cl}^-$ ,  $\text{NO}_3^-$ ,  $\text{SO}_4^{2-}$ ), nutrients (total nitrogen, total phosphorus,  $\text{NO}_3^-/\text{NO}_2^-$ -N,  $\text{NH}_3$ ), turbidity, color, pH, conductivity, dissolved oxygen, dissolved organic and inorganic carbon. They were analyzed following standard methods (US EPA, 1989). Physical habitat characteristics were summarized (mean, standard deviation) from 20-100 measurements per site (Kaufmann et al., 1999). Physical habitat was measured

along the sampling reach to determine channel dimensions and gradient, substrate size and type, habitat complexity and cover, riparian vegetation cover and structure, anthropogenic alterations and channel-riparian interaction (Kaufmann, 2006). The landscape component of WEMAP included soils, topography, climate, vegetation and land use. Watershed boundaries were delineated from 1:24,000-scale USGS topographic maps and USGS's 30-m digital elevation models (DEMs). They were used to calculate watershed area, slope and elevation. Watershed land use/cover were characterized from the USGS 1992 National Land Cover Dataset.

## **Chapter 1: The importance of rare taxa in diatom-based stream bioassessment.**

### **Introduction**

Bioassessment uses the structure and function of biotic assemblages to detect and evaluate environmental impairment (Karr, 1999; Barbour et al., 1999). Every member of the assemblage provides information about their abiotic environment and biotic interactions. Unlike common taxa which contribute most of the species abundance in an assemblage, rare taxa may constitute the largest component of species richness in an assemblage (Marchant et al., 1997; Potapova and Charles 2002, 2003; Hessen and Walseng, 2008). For example, the dominant diatom species in an assemblage may account for up to 80% of a sample's relative abundance (Weilhoefer and Pan, 2008), while 70% of all diatoms (1028 out of 1461 taxa) may be considered rare and excluded from further analysis (Potapova and Charles, 2002). Despite their recognized contribution to species richness, rare taxa remain a topic of ongoing debate in bioassessment.

The importance of rare taxa in bioassessment is still unclear. Several studies came to contradictory conclusions about their responses to impairment. Hawkins et al. (2000) found that rare macroinvertebrates increased in abundance at impacted stream sites compared to reference ones, while Cao et al. (1998) concluded the opposite. One possible explanation for these contradictory results may be different environmental regulators of common and rare taxa. For instance, Faith and Norris (1989) compared correlations between 24 physico-chemical variables and ordinations of common and

rare macroinvertebrates and found higher number of significant correlations for rare taxa (17) than for common taxa (10). They suggested that the distribution and abundance of rare taxa might be determined by an additional set of environmental variables (e. g., velocity, conductivity, and total dissolved solids) compared to common taxa (e. g., temperature, stream order, particle size, and water chemistry). Another possible explanation for these contradictory results may be sampling limitations, because most bioassessment programs are not designed with rare taxa in mind and rare taxa may only add noise to data analyses (Marchant, 1999, 2002; Hawkins et al., 2000; Van Sickle et al., 2007). Most studies that examined the effects of rare species exclusion on the detection of major environmental gradients found that common species recovered these gradients well (Cao et al., 1998; Van Sickle et al., 2007; Hawkins et al., 2000; Nijboer and Schmidt-Kloiber, 2004). Marchant (1999, 2002) suggested that the information contained in rare taxa may already be provided by common species which are better sampled and quantified. A third possible explanation for the contradictory results of rare taxa in bioassessment may be their heterogeneous responses to the same set of environmental variables, because different subsets of rare taxa may reveal different environmental patterns (Faith and Norris, 1989). For example, Nijboer and Schmidt-Kloiber (2004) found that when macroinvertebrate taxa with low relative abundances were excluded from the calculation of a multimetric index of ecological quality, sites were classified in higher ecological quality class than when using all taxa, but when taxa with narrow distribution ranges were excluded the results were the opposite. They suggested that

taxa with low relative abundances were associated with human disturbance, while taxa with narrow distribution ranges were indicators of reference conditions in the Netherlands.

Compared to macroinvertebrates, the potential bioassessment value of rare diatoms in streams has not been well studied (Potapova and Charles, 2004; Lavoie et al., 2009). Potapova and Charles (2004) found that the number and abundance of rare diatoms in US rivers had negative correlation coefficients and therefore decreased along gradients of elevation and human disturbance (nutrient concentrations and percent of urban and agricultural land use in the watershed). Lavoie et al. (2009) concluded that ordinations of epilithic diatom assemblages from Eastern Canada differed depending on the criteria for taxa exclusion. They found that excluding 40% of taxa based on relative abundance in a sample ( $\leq 2\%$ ) did not affect the results, but excluding taxa based on their frequency of occurrence ( $\leq 3$  sites) greatly affected the structure of the ordination. Both of these studies used different definitions of rarity (either abundance or occurrence) which makes their comparison difficult. There is no universal definition of rarity, but most commonly it is defined in terms of abundance (low or high) and/or occurrence (low or high) (Gaston, 1994; Cao et al., 1998; Hawkins et al., 2000; Potapova and Charles, 2002, 2004; Nijboer and Schmidt-Kloiber, 2004; Carlisle et al., 2008). The combinations among them result in three definitions of rare species: both are low (satellite taxa *sensu* Hanski (1982)), low occurrence and high abundance (urban taxa *sensu* Söderström (1989)), and high occurrence and low abundance (rural taxa *sensu* Söderström (1989)). The taxa in each

one of these rarity categories are characterized with specific biological features like dispersal ability and population stability (Söderström, 1989), and may therefore respond to different sets of environmental factors. Rare taxa are assumed to have narrow ecological niches, to be sensitive to pollution and to be potential valuable bioindicators (Potapova and Charles, 2002). However, it is still unknown what environmental variables determine rare taxa richness in a site and whether it can be used in bioassessment as an indicator of specific environmental conditions.

Diatom rarity is most commonly associated with dispersal ability. Diatoms may be dispersed by water, biota (e.g., insects, birds, fish and mammals), humans and air (Kristiansen, 1996). Some researchers insist that the majority of diatom species are endemic (Kociolek and Spaulding, 2000; Vyverman et al., 2007; Vanormelingen et al., 2008) while others recognize their cosmopolitan distribution (Kristiansen, 1996; Finlay et al., 2002). One group of scientists promotes the phylogenetic view of biogeography according to which ancient species are widely distributed while recently evolved species have restricted distributions because they have not had enough time to spread (Theriot, 1992). For example, *Stephanodiscus yellowstonensis* Theriot and Stoermer is considered endemic to Yellowstone Lake, Wyoming, because it has been found in geologically young sediments (Theriot, 1992). Yet, studies from ancient lakes such as Baikal and Ochrid have concluded that their unique diatom floras are due to their isolation and ancient origin (Skvortzow and Meyer, 1928; Skvortzow, 1937; Jurilj, 1954). Another group of scientists argues that worldwide distribution is evidence for good dispersal abilities (Round, 1981) and more than 900 diatom genera



are considered to be cosmopolitan (Fourtanier and Kociolek, 2003). Such cosmopolitan distributions are not uncommon for microorganisms which made Baas Becking (1934) hypothesize that “everything is everywhere, but the environment selects”. The success of a species in a local habitat may largely be determined by a continuum of spatial (watershed, valley, channel unit and microhabitat) and temporal factors which function as a set of environmental filters at different hierarchical landscape scales (Tonn, 1990; Poff, 1997; Stevenson, 1997). Each filter constrains the available environmental conditions and reduces the number of species able to pass through it. For example, widely distributed taxa may be able to disperse well, pass through the majority of environmental filters and either establish themselves under favorable conditions with high local abundances or strive to survive. As pointed out by Stevenson (1997), algal assemblages in riffles may differ from pools due to faster current velocity and supply rates of nutrients which are determined by geology and land use. For example, some species have high affinity for specific habitats which restricts their occurrence in others (e. g., *Gomphonema parvulum* (Kützing) Kützing occurred only on stone substrates, while *Cocconeis placentula* Ehrenberg occurred only on plants, Soininen and Eloranta, 2004).

The study of rarity may depend on, among other things, the sampling protocol and scale, sample size and survey intensity (Cao et al., 2001). The US Environmental Protection Agency’s Western Environmental Monitoring and Assessment Program (WEMAP) provides a great opportunity to enhance our understanding about the importance of rare taxa in diatom-based bioassessment. In addition to its large spatial

coverage and probability-based sampling design (Herlihy et al., 1998) which allows conclusions to be made about the whole study area, WEMAP employs a unified sampling and processing protocol to permit direct comparisons between samples from vastly distant and diverse geographical areas. The multi-transect composited samples per stream reach are designed to capture habitat heterogeneity and its associated species diversity. The WEMAP dataset presents a unique chance to study rarity in diatom communities whose distribution patterns, biogeography and diversity are not well known (Martiny et al., 2006).

The major objective of this chapter was to evaluate the potential bioassessment value of rare diatoms in streams/rivers. Specifically, I wanted to characterize what environmental factors may be affecting the number of rare taxa in Western US streams and rivers. Next, I wanted to test if the number of rare taxa were significantly different between streams with contrasting human disturbance. I anticipated that if rare taxa were truly sensitive to pollution, there will be fewer of them at impacted sites compared to reference ones. In addition, I wanted to assess if selected rare taxa have narrower ecological niches than some common taxa.

## **Materials and Methods**

*Dataset* - The site selection and field sampling were conducted as part of the WEMAP protocols outlined in the *Background Information* section. Water chemistry, physical habitat, landscape characteristics and diatom data were collected from perennial streams and rivers in 12 Western states. A subset of 1037 streams and rivers

with 600 diatom counts, required by standard counting methods (Charles et al., 2002), was used for this analysis.

*Sample collection* - One periphyton sample was collected at ¼, ½ or ¾ of the distance to the stream bank for each of the 11 transects. A surface area of 12-cm<sup>2</sup> was scraped with a toothbrush from rock/wood substrates (erosional habitat); soft sediments were vacuumed with a 60-mL syringe (depositional habitat). All 11 samples per reach were combined into one composite sample which was preserved with 37% formalin.

*Diatom analysis* - A small amount of each composite diatom sample was acid-digested (Patrick and Reimer, 1966), rinsed with distilled water and mounted on a microscope slide in ZRAX<sup>®</sup> (refraction index 1.70) or NAPHRAX<sup>®</sup> (refraction index 1.73) mounting medium. A minimum of 600 diatom valves were counted and identified to the lowest possible taxonomic level (mainly species) at 1000X under Leica DMLB microscopes with Differential Interference Contrast (DIC) (Leica Microsystems, Inc.). The diatom taxonomy followed primarily Krammer and Lange-Bertalot (1986, 1988, 1991a, b) and Patrick and Reimer (1966, 1975).

*Data screening* – Diatom counts were represented as a percentage of total species counts per site (relative abundance). Some taxa (4) were identified to their closest taxonomic identities with some reservations by the original analysts. In this study, these taxa were merged with their closest relative (e. g., *Fragilaria* cf. *vaucheriae* (Kützing) Petersen with *Fragilaria vaucheriae* (Kützing) Petersen). The varieties of *Cocconeis placentula* Ehrenberg were included in the nominate variety

because the lower part of their cells could not be consistently distinguished from the upper one. All unknown taxa (123, mean relative abundance 2.91%, range 0.16-59.50%) were removed from the analysis to avoid artificial overestimation of rarity. This procedure led to the deletion of 30 sites due to low diatom counts. As a result, sites with more than 540 valve counts (> 90% total relative abundance) were retained for the subsequent analysis. Another 20 sites were excluded because of missing values for more than ten environmental variables. The final dataset consisted of 987 sites (Fig. 1.1).

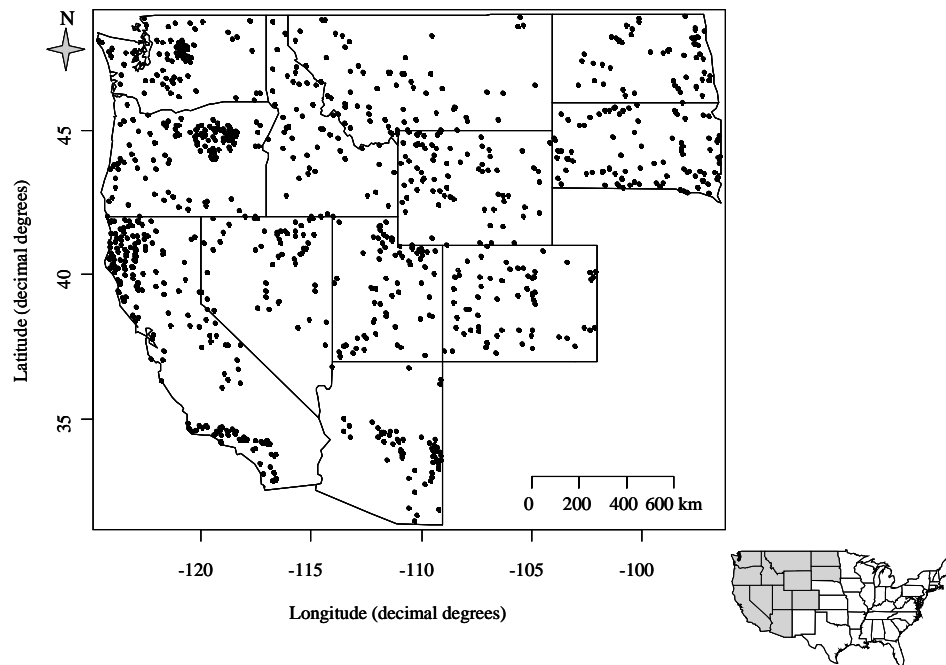


Figure 1.1. Maps of USA and 12 Western states with locations of sampling sites (n=987).

*Rarity metrics* - Taxa were delineated by local relative abundance (site-specific metrics) and by combinations of occurrence (number of sites) and mean relative abundance at the sites where present (regional metrics). Relative abundance of 1%

which is a widely used in diatom assessment (Potapova and Charles, 2002, 2003) was selected as the cutting line between rare and common taxa. Species occurrence at 5 or less sites and at 30 or more sites were the dividing lines for restricted and wide distributions, respectively. Two site-specific rarity metrics were employed: rare 1% (taxon with relative abundance  $\leq 1\%$  in a sample) and max 1% (taxon with relative abundance  $\leq 1\%$  from the relative abundance of the most abundant taxon at a site). Regional metrics were represented as the number of species in each of the following classes: urban, rural (*sensu* Söderström, 1989), satellite and core (*sensu* Hanski, 1982). The first three classes are all rare in some sense, whereas the last one represents common taxa (Söderström and During, 2005). This terminology was selected for ease of data interpretation. Each species was assigned to a rarity class based on the criteria in Table 1. 1. Two site-specific (rare 1% and max 1%) and three regional (satellite, urban and rural) rarity metrics, as well as one core metric were summarized for the whole dataset, the three ecoregions (Mountains, Plains and Xeric), reference and impacted sites. Site-specific rarity metrics had no identities, because the same species could be classified as rare in one site and common in another. Therefore, they were not examined or discussed in great detail.

Table 1.1. Regional rarity classes.

Abundance (mean % RA)	Occurrence (number of sites)	
	$\leq 5$	$> 30$
$\leq 1\%$	satellite	rural
$> 1\%$	urban	core

Satellite and core definitions *sensu* Hanski (1982)

Urban and rural definitions *sensu* Söderström (1989)

*Rarity predictors* - To characterize environmental factors which may be affecting the number of rare taxa in Western US streams and rivers, I used regression tree (RT) analysis (Breiman et al., 1984). This analysis is used for finding important environmental predictors and their interactions in terms of the response variable (rarity metric). Unlike multiple regression, RT algorithm can handle missing values and does not require any assumptions about the data. A regression tree is constructed after a series of binary decisions are made based on the predictors. It splits the dataset into smaller and smaller groups, where each split depends on a single predictor variable. The growth of the RT starts with a split of the dataset (*root node*) into two groups (nodes) with maximum homogeneity in the values of their response variable (rarity metric) (Fig. 1. 2a). Homogeneity is measured by deviance calculated as the sum of squared deviation of the response variable from its mean (see the formula in Fig. 1. 2b). Then each one of the two new groups is split (*node*) again into two new groups based on the next predictor variable with the highest discriminatory power and the process is repeated again until the terminal nodes (*leaves*) are too small to be split (Fig. 1. 2b). Each leaf provides the predicted mean value for the response variable (rarity metric), the number of objects in the group (sites) and decision rules (predictors) that define it.

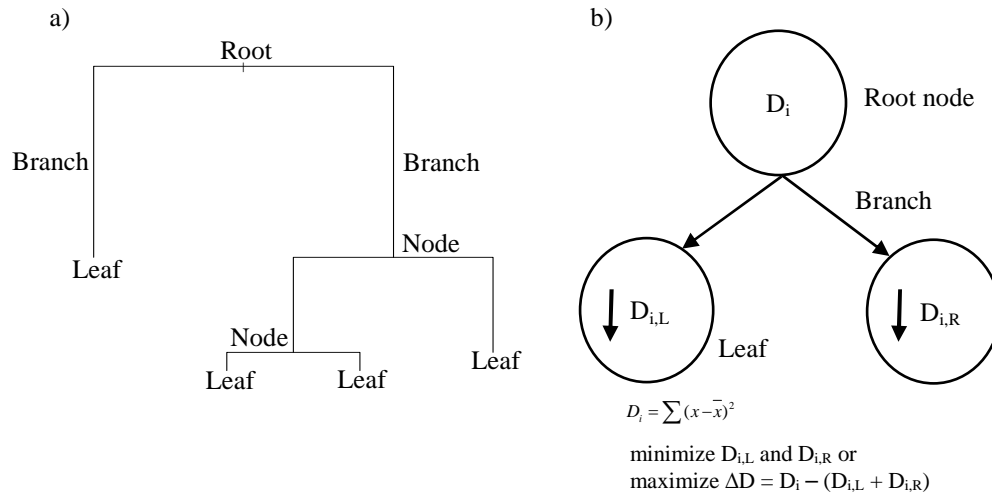
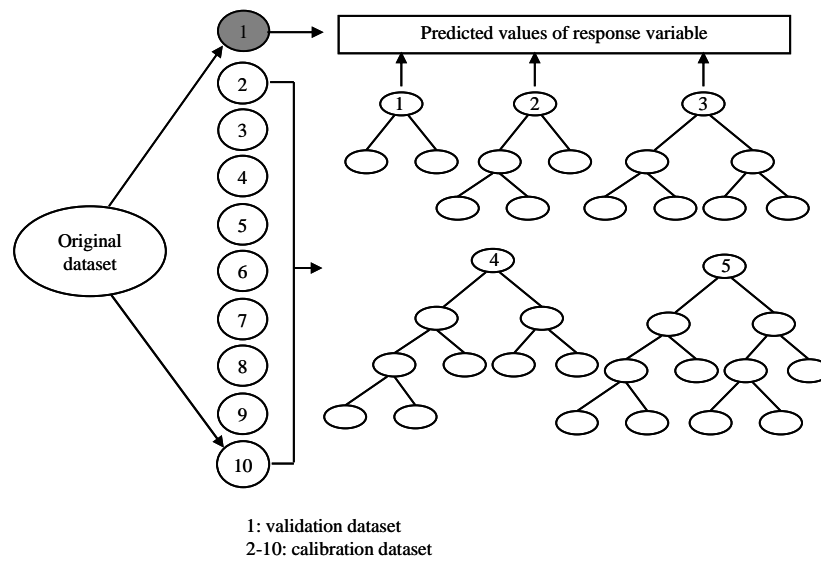


Figure 1.2. Regression tree building: a) an example of the terminology of a regression tree; b) “growing” a regression tree,  $D_i$  – deviance in the response variable at the root of the tree,  $D_{i,L}$  – deviance in the response variable in the left new group,  $D_{i,R}$  – deviance in the response variable in the right new group,  $i$  – response variable.

The resulting full tree may explain the variability in the response variable well but may have limited predictive power due to its complexity (i.e. number of splits). Therefore, cross-validation is used to trim the tree and produce a less complex final tree. Cross-validation is done by randomly partitioning the original dataset in two subsets, one containing 90% of the sites for model building (calibration dataset) and another with 10% for testing (validation dataset). This results in many tree models with different sizes (number of splits). The difference between the observed and predicted values of the response variable in the validation dataset is used to calculate the model prediction error for each tree size (Fig. 1. 3a). Cross-validation is repeated 10 times, generating unique sets of prediction errors for the different tree sizes. The prediction errors from each cross-validation run are used to calculate the relative

prediction error for the different tree sizes. The size of the final tree model is determined by plotting the mean relative prediction errors against the number of splits (tree sizes) and selecting the one with the lowest relative prediction error (Fig. 1. 3b). To reduce the large amount of natural variability in the dataset, it was split by ecoregion and reference condition, and the regression tree analysis was repeated.

a)





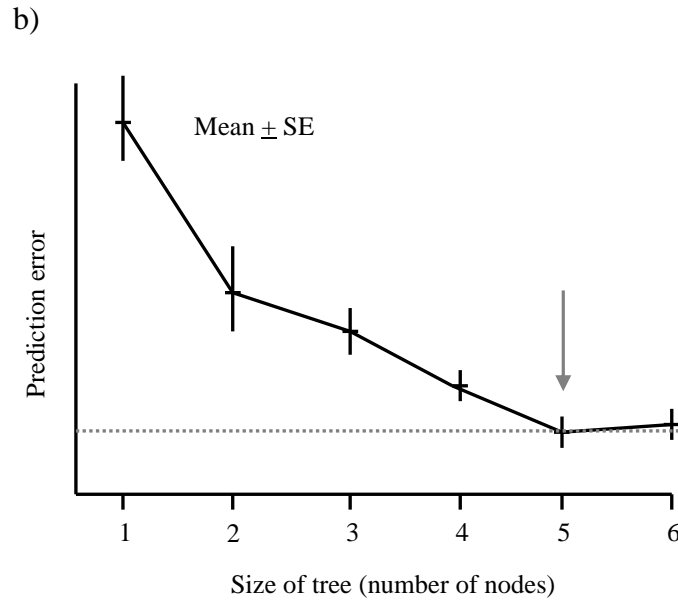


Figure 1.3. Regression tree cross-validation with multiple trees of different sizes (a) and selection of final regression tree with the lowest mean prediction error (b).

*Reference vs. impacted sites* –WEMAP sites were classified by EPA as least disturbed (reference) and most disturbed (impacted) within each of the aggregated level III ecoregions (i.e. Mountains, Plains, Xeric) (Omernik, 1987; US EPA, 1996). EPA used four sets of criteria to categorize sites as reference (Whittier et al., 2007). First, the natural variability within each ecoregion was captured by finding the most important natural gradient (e. g., stream size, elevation and slope in the Mountains and Xeric, and longitude in the Plains). Second, for each ecoregion, 11 or 12 anthropogenic stressors of relevant importance to water quality (total phosphorus, total nitrogen, chloride, sulfates, pH, turbidity), physical habitat (riparian disturbance, % fine particles in the substrate or relative bed stability, riparian vegetation) and

watershed disturbance (% urban land, % agricultural land, human population density and road density) were selected. Third, each stressor was plotted against the natural variable with which it was most highly correlated, and 15% of the sites with the lowest and 15% of the sites with the highest values along the natural gradient were scored as least and most disturbed, respectively. For example, Mountain sites with  $>20 \mu\text{g/L}$  total phosphorus and  $> 750 \mu\text{g/L}$  total nitrogen were not considered as being in reference condition, but sites with the same concentrations in the Plains were (Waite et al., 2000; Herlihy et al., 2006; Whittier et al., 2006). Fourth, the number of times a site fell within the least disturbed category was summed and if it was higher than the number of times the site fell within the impacted category, the site was classified as reference. Otherwise, it was classified as impacted. In addition, EPA personnel conducted a number of discussions to select the final set of least disturbed and most disturbed sites within each ecoregion. To test if the number of rare taxa were significantly different between reference and impacted sites, they were compared with t-tests ( $\alpha=0.05$ ). This comparison was done for the whole dataset, three ecoregions (Mountains, Plains and Xeric), two stream order classes ( $1^{\text{st}}$ - $3^{\text{rd}}$  and  $> 3^{\text{rd}}$ ; Strahler, 1957) and two successional stages (early and late) within each ecoregion. Sites were split based on stream order to create groups with similar hydrologic characteristics and to evaluate whether they affect rarity metrics. Successional stages were used to evaluate whether the number of rare taxa may be related to community development. I anticipated that late successional stages will have more rare taxa as a result of their higher species richness. Sites were grouped in two categories as being in early or late

succession based on the sum of relative abundances of species with ruderal (R) strategy (Biggs et al., 1998). These species are early colonizers of denuded habitats and reach high abundances at the start of succession. Sites with more than 50% relative abundance of R strategy species were classified as being in early successional stages, while sites with less than 10% of R strategy species were considered in late successional stages.

*Ecological niches* - To estimate the width of ecological niches of selected rare and common taxa, I used the weighted averaging approach (ter Braak and Looman, 1986). This method is based on the idea that a single environmental variable determines the species composition at a site and that each species has a bell-shaped response curve with regard to this variable. The species response curve is summarized as species optimum (indicator value, mode of the curve) and tolerance (ecological amplitude). Species optimum is calculated by averaging the environmental variable over the samples where the species is present weighted by the species abundance in the sites. Tolerance is calculated as the weighted standard deviation of species abundance in the sites. Weighted average (WA) values were calculated for all important variables from the regression trees and total disturbance in the watershed. WA is one of the most widely used methods to infer species niches, it is straightforward to apply and it provides reasonable estimators of species optima, if they have few occurrences and a small tolerance (ter Braak and Looman, 1986). The coefficient of variation (weighted average tolerance divided by weighted average optimum) was used to make the results more comparable and avoid the influence of

taxa with high means. The coefficient of variation for the five species with the highest occurrence and abundance values in each of the common (core) and rare (rural, urban and satellite) regional classes are presented.

All data analyses were performed with R language and environment for statistical computing (R Development Core Team, 2008).

## **Results**

Diatom assemblages in Western US streams were species rich. A total of 950 taxa were identified in the 987 samples for this study. Taxa richness varied widely among the samples. The mean number of taxa in a sample was 42 (range 11-95) (Table 1. 2). On average, sites in the Plains ecoregion had the highest taxa richness (54) while Mountain and reference sites had the lowest taxa richness, 38 and 37 taxa, respectively. The most taxa rich genera were *Navicula* (136 taxa), *Nitzschia* (109) and *Gomphonema* (54). On average, the highest number of taxa per site was recorded for core species (30), rare 1% (27) and rural species (7) (Table 1. 2). The largest variation in the number of taxa was exhibited by rare 1% (3-69), max 1% (0-44) and core (4-62) metrics. The lowest numbers of taxa were recorded for rare 1%, rural and core species from Mountain sites, followed by sites in the Xeric and Plains ecoregions.

Table 1.2. Sample characteristics and number of taxa in each rarity class with mean (minimum and maximum) values for the whole dataset (West), three ecoregions (MT-Mountains, PL-Plains, XE-Xeric), reference and impacted sites.

	<b>West (n=987)</b>	<b>MT (n=571)</b>	<b>PL (n=191)</b>	<b>XE (n=225)</b>	<b>Reference (n=194)</b>	<b>Impacted (n=165)</b>
No. taxa	950	795	589	613	584	605
Sample richness	42 (11-95)	38 (11-83)	54 (14-95)	42 (15-79)	37 (11-87)	46 (17-86)
Species occurrence (no. sites)	44 (1-893)	28 (1-546)	17 (1-166.00)	15 (1-201)	12 (1-179)	12 (1-142)
Species mean RA	1.11 (0.16-16.86)	1.04 (0.17-20.21)	1.10 (0.17-26.11)	1.34 (0.16-41.74)	1.20 (0.17-27.32)	1.31 (0.16-16.63)
<b>Rarity classes</b>						
Rare 1%	27 (3-69)	24 (3-59)	35 (8-69)	26 (9-53)	23 (7-66)	29 (10-59)
Max 1%	6 (0-44)	7(0-44)	4 (0-34)	6 (0-41)	7 (0-44)	5 (0-41)
Satellite	< 1 (0-8)	< 1 (0-8)	< 1 (0-5)	< 1 (0-7)	< 1 (0-8)	< 1 (0-7)
Urban	< 1 (0-4)	< 1 (0-4)	< 1 (0-3)	< 1 (0-2)	< 1 (0-2)	< 1 (0-2)
Rural	7 (0-23)	6 (0-23)	10 (0-22)	7 (0-19)	6 (0-19)	8 (1-23)
Core	30 (4-62)	28 (4-53)	37 (11-62)	30 (7-56)	27 (7-60)	32 (10-52)

*Rarity metrics* - Diatom occurrence and abundance showed several patterns in the Western US streams (Fig. 1. 4). Most noticeably, only a few taxa were both locally abundant and widely distributed. The two taxa with the widest distribution ranges were *Achnantheidium minutissimum* (Kützinger) Czarnecki (851 sites, 86.22% of sites) and *Cocconeis placentula* Ehrenberg and vars. (893 sites, 90.48% of sites). Their mean relative abundances were 16.86% (range 0.17 - 97.19%) and 8.66% (range 0.16 - 88.33%), respectively. In contrast, 205 taxa (21.58%) occurred at only one site. However, their mean relative abundances varied substantially (Fig. 1. 4). More than two thirds of the taxa (654 taxa, 68.84%) occurred with  $\leq 1\%$  relative abundance. Urban taxa, with few exceptions, were relatively small ( $< 50 \mu\text{m}$ ), but with very diverse morphology and ecology. The most taxa-rich rarity class was satellite (362 taxa), followed by rural (99 taxa) and urban (70 taxa) (Table 1. 3). The satellite class comprised a large number of taxa in the genera *Navicula* (59), *Nitzschia* (29) and *Eunotia* (27). The most taxa-rich genera in the urban class were *Gomphonema* (9), *Navicula* (7) and *Eunotia* (6). Most of the taxa in the rural rarity class belonged to the genera *Navicula* (19) and *Nitzschia* (15). They occurred on average at 44 sites (range 1-893) with a mean of 1.11% relative abundance (range 0.16-16.86%).

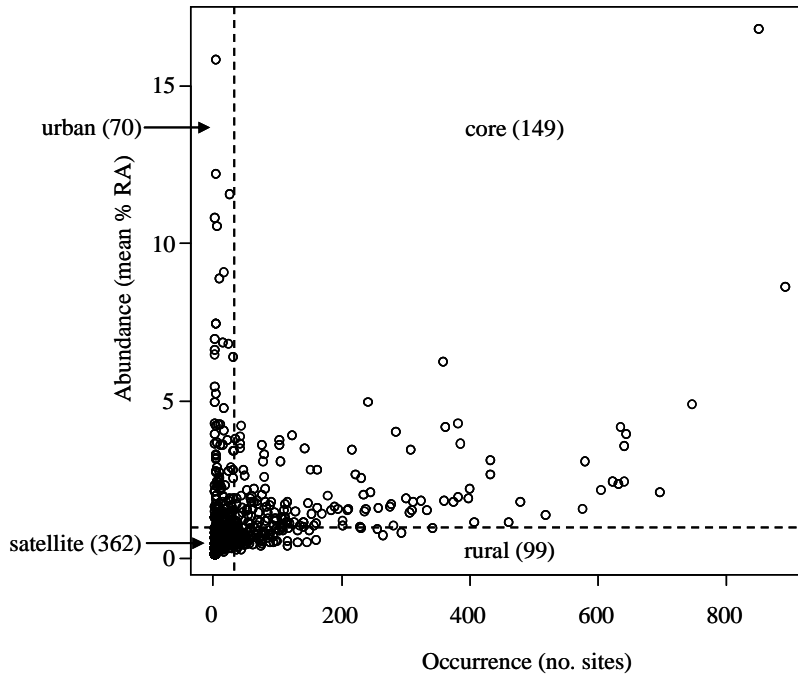


Figure 1.4. Plot of species occurrence and mean % relative abundance (RA) in the Western US streams/ivers. Each point represents a diatom taxon. Dashed lines define boundaries between rarity classes with number of taxa in them (numbers in parentheses).

Table 1.3. Number of taxa and mean relative abundance (minimum-maximum) for regional rarity classes.

Rarity classes	No. taxa (n=950)	No. genera (n=103)	Mean RA (Min-Max)
Satellite	362	79	0.40 (0.16-1.00)
Urban	70	37	3.05 (1.04-15.89)
Rural	99	39	0.70 (0.29-1.00)
Core	149	52	2.19 (1.01-16.86)

*Rarity predictors* - Environmental variables explained up to one third of the variation in the rarity metrics in the Western US. Regression tree (RT) models were developed for three (rare 1%, max 1% and rural) of the five rarity metrics (Table 1. 4). For the other two rarity metrics (satellite and urban), cross-validation indicated that none of the predictor variables could explain the variation in their values better than

by pure chance. The number of nodes, a measure of the tree size, for the RT models ranged between one and ten. The variables most often associated with the variation in the metrics were substrate as percent of sand and fines, distance to ocean and watershed slope. The goodness-of-fit of the RT models ranged between 0.08 and 0.36. The best models (the ones with the highest  $r^2$ ) were for the core metric (10 nodes,  $r^2 = 0.36$ ), rural (6 nodes,  $r^2 = 0.32$ ) and rare 1% (4 nodes,  $r^2 = 0.30$ ).



Table 1.4. Summary of the regression tree (RT) models used to associate variation in rarity metrics with environmental features (predictors) for the whole dataset (West), two ecoregions (MT-Mountains and PL-Plains) and reference sites. No models were developed for Xeric ecoregion and impacted sites. Number of nodes and variables are measures of the complexity of the models.  $r^2$  measures the strength of association between the metric and the predictors. Metric descriptions and abbreviations are in text and Table 1.1. Variable abbreviation and summaries are in the Appendix.

Rarity metrics	West			MT			PL			Reference		
	$r^2$	No. splits	Variables	$r^2$	No. splits	Variables	$r^2$	No. splits	Variables	$r^2$	No. splits	Variables
Rare 1%	0.30	4	PCT_SAFN, SLOPMEAN, XDEPTH, KM_SEA	0.25	3	PCT_SAFN, KM_SEA, PCT_FAST	-	-	-	0.42	2	PCT_SAFN, TURB
Max 1%	0.08	1	COND	0.21	4	COND, BAR_TOT, TURB, SIO2	-	-	-	0.50	6	COND, PCT_FAST, WSAREA, BAR_TOT, PTL, SIO2
Satellite	-	-	-	-	-	-	-	-	-	-	-	-
Urban	-	-	-	-	-	-	-	-	-	-	-	-
Rural	0.32	6	PCT_SAFN, KM_SEA, LAT_DD, NO3, SLOPMEAN, WSAREA	0.25	2	PCT_SAFN, KM_SEA	0.25	2	WSAREA, PCT_SAFN	0.56	4	PCT_SAFN, TURB, PRECIP_M, PCT_SAFN
Core	0.36	10	PRECIP_M, PRECIP_M, WSAREA, XCDENBK, KM_SEA, TURB, COND, TURB, XCDENBK, PCT_FAST	0.17	2	PRECIP_M, PRECIP_M	0.33	3	COND, KM_SEA, XCDENBK	0.56	5	TURB, NO3, PCT_SAFN, LAT_DD, PRECIP_M

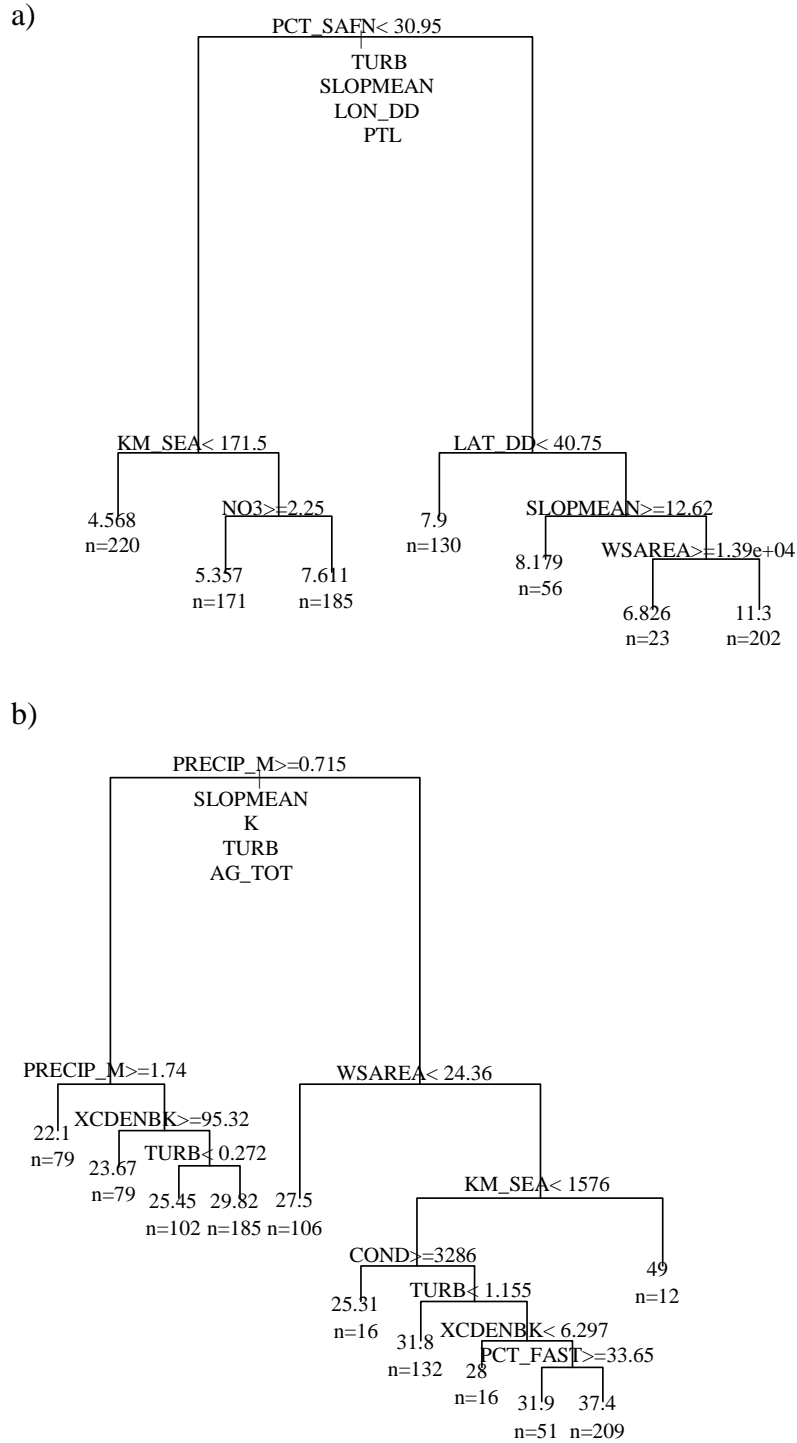


Figure 1.5. Final regression trees for rural (a,  $r^2=0.32$ ) and core (b,  $r^2=0.36$ ) metrics. Names and numbers above each node indicate the variable and its value used for the split. Sites which meet the node criteria are split to the left. Names below the first split indicate alternative variables. Numbers at the leaves specify mean predicted values for the response variable (rural and core metrics) and the number of sites (n). Variable descriptions are in the Appendix.

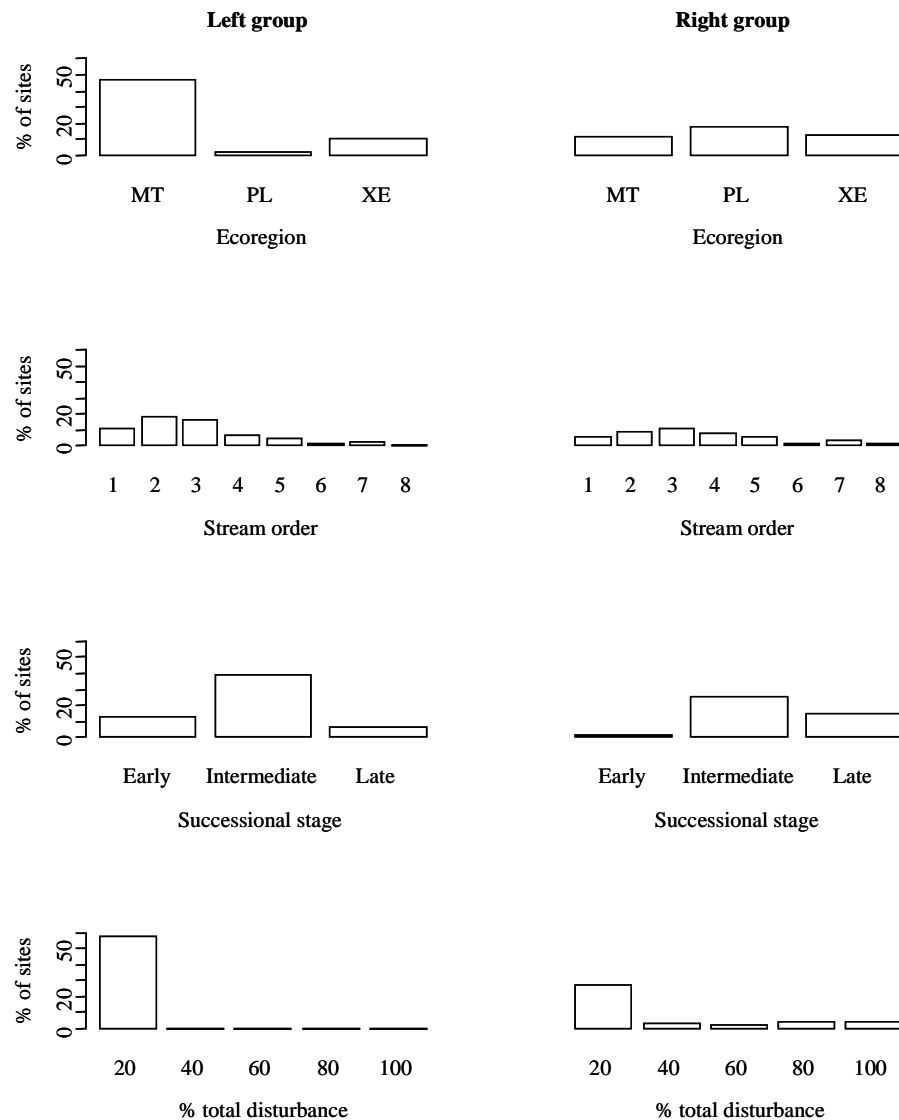


Figure 1.6. Site summaries for the left and the right nodes on Fig. 1.5a.

In addition to the above “best” model, several alternative RT models also performed well. They included a few other variables which could almost equally well split the data into the two groups and define a similar longitudinal gradient. These

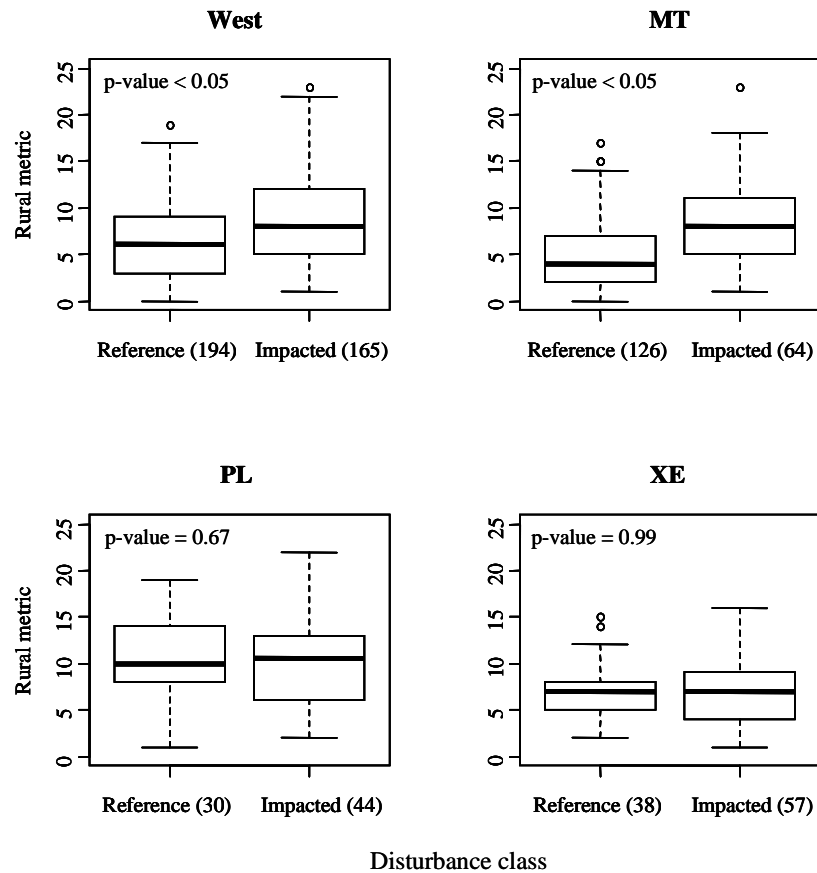
variables were turbidity, mean watershed slope, longitude and total phosphorus concentrations (Fig. 1.5a).

In contrast, the regression tree for the core metric had ten splits and the most important variables were precipitation and watershed area (Fig. 1.5b). The number of core taxa at sites with higher precipitation ( $\geq 0.72$  m, left node) was determined by bank side canopy density and turbidity. The number of core taxa at sites with less precipitation ( $< 0.72$  m, right node) and bigger watershed area was influenced by distance to ocean, water chemistry (conductivity and turbidity) and physical habitat characteristics (bank side canopy density and fast water habitat). Only when the data was split into reference and impacted sites did the regression trees explain much more of the variation in the rarity metrics. The biggest change was recorded for the rural metric which increased at reference sites to 56% (Table 1.4).

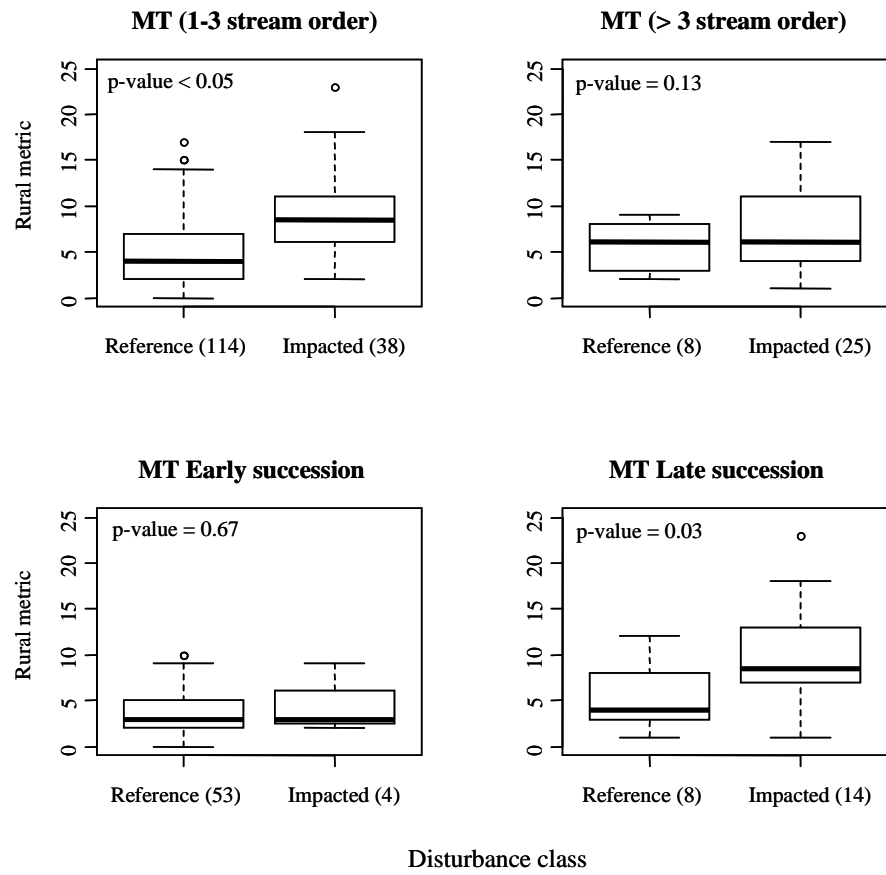
*Reference vs. impacted sites* – Since no models were developed for the satellite and urban metrics, their controlling variables were unknown and these two metrics were not used for the comparison of reference and impacted sites. In addition, they had very narrow ranges (0-4 for the urban metric and 0-8 for the satellite one) which would not distinguish reference from impacted sites. Rural species richness at impacted sites was significantly higher than that of reference sites in Western streams/rivers ( $p < 0.05$ ) and Mountain sites ( $p < 0.05$ ) (Fig. 1.7a, b). Neither of the other two ecoregions (Plains and Xeric) exhibited significant differences in the rural metric between reference and impacted sites ( $p = 0.67$  and  $p = 0.99$ , respectively) (Fig. 1.7c, d). The mean number of rural taxa at reference sites in the Mountains (5.12) was lower

than in the Plains and Xeric ecoregions (10.57 and 7.08, respectively). When the reference and impacted sites in each ecoregion were further split by stream order and successional stage, the number of rare taxa were significantly different only for low order streams ( $p < 0.05$ ) and late successional stages ( $p = 0.03$ ) in the Mountains.

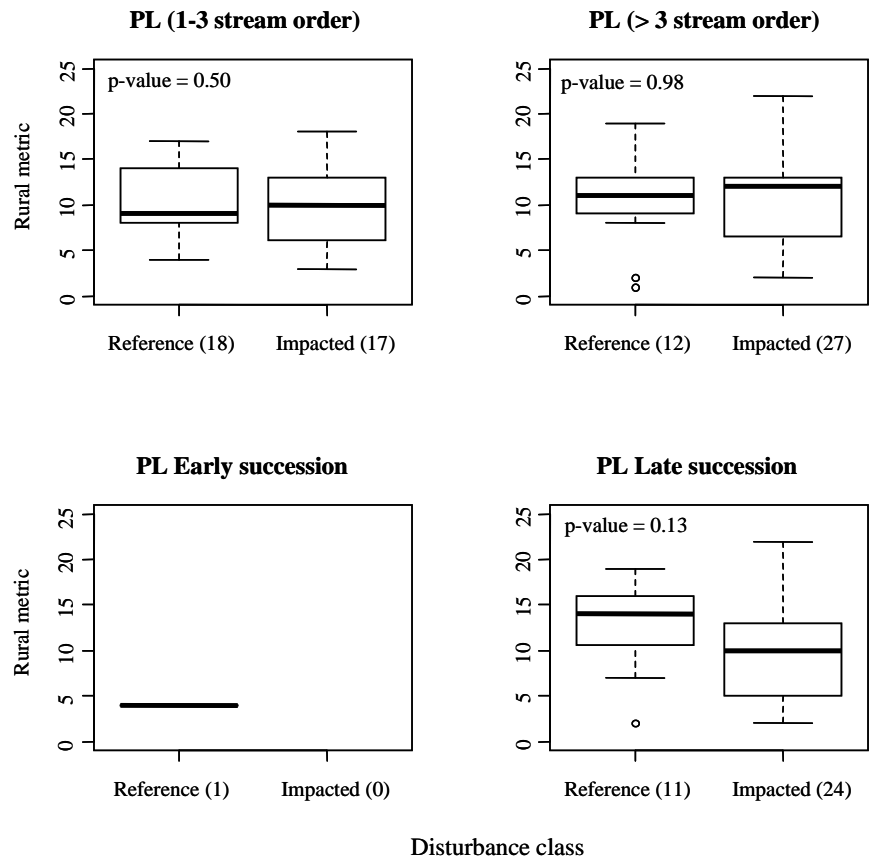
a)



b)



c)



d)

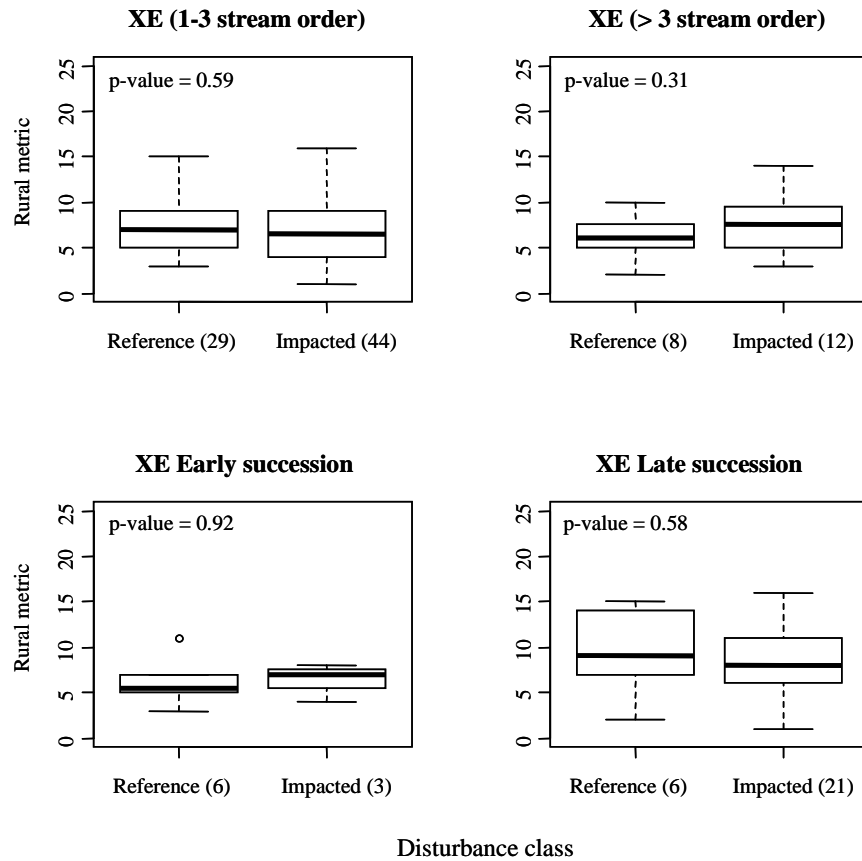


Figure 1.7. Rural metric at reference and impacted sites in the dataset and the three ecoregions (MT, Mountains; PL, Plains; XE, Xeric) (a) and in each of the ecoregions divided by stream size and diatom successional stage: MT (b), PL (c) and XE (d). Numbers in parenthesis show the number of sites.

*Ecological niches* - Rare and common taxa did not exhibit pronounced differences in the width of their ecological niches (Table 1.5). Their weighted averaging optima and tolerances overlapped. No clearer pattern was revealed by the coefficient of variation. The only pronounced difference was higher values for the common taxa compared to the rural taxa for the variable substrate as percent of sand and fines in the West and total phosphorus concentrations in the Mountains (Fig. 1.8).



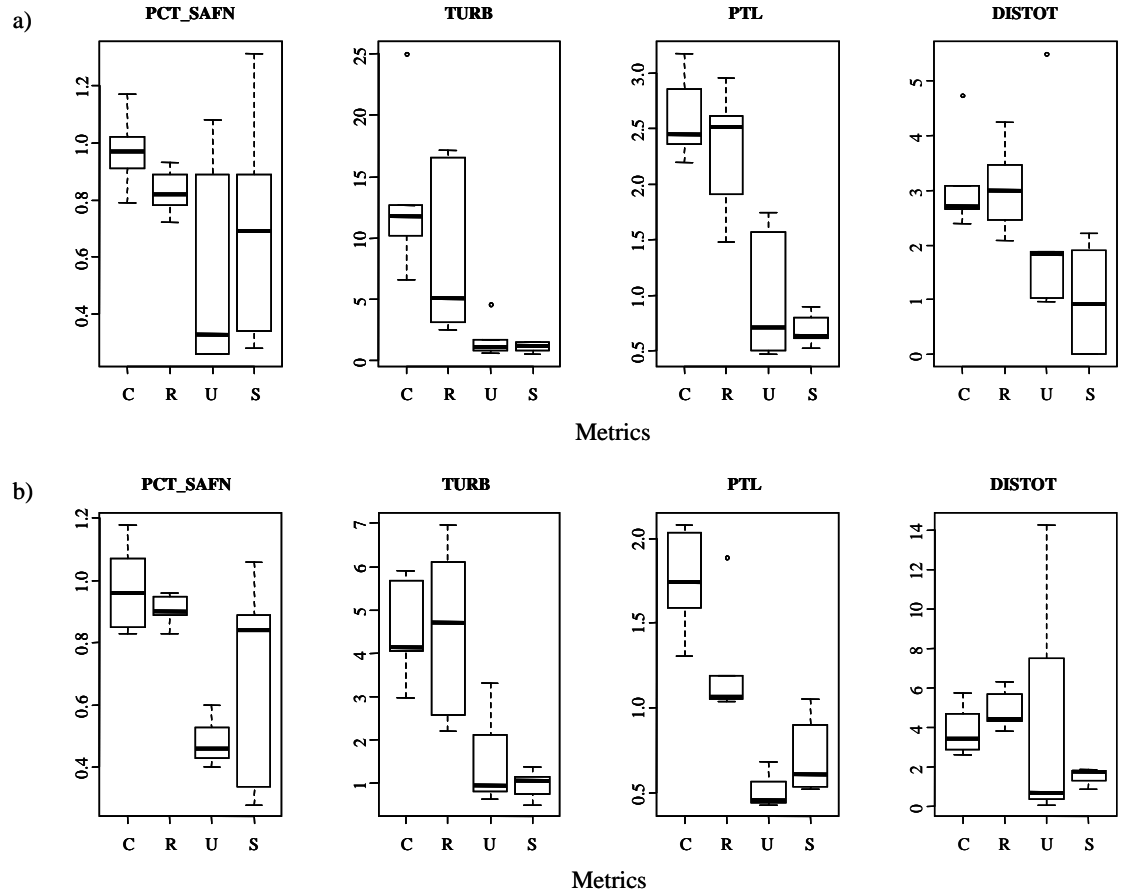


Figure 1.8. Boxplots of coefficient of variation (weighted average tolerance/ weighted average optima) for five core (C), rural (R), urban (U) and satellite (S) taxa in the whole dataset (a) and in the MT ecoregion (b). Their names and summaries are provided in Table 1.5. Metric descriptions are in Table 1.1. Variable abbreviations and summaries are in the Appendix.

Table 1.5. Weighted average values, optima (tolerances), number of sites and mean relative abundances for selected rare and common taxa. Variable abbreviation and summaries are in the Appendix.

Common taxa	n	Mean RA (min-max)	PCT_SAFN (%)	TURB (NTU)	DISTOT (%)	PTL (ug/L)
<i>Achnanthidium minutissimum</i> (Kützing) Czarnecki	851	16.86 (0.17-97.19)	20 (23)	6 (138)	2 (9)	22 (68)
<i>Cocconeis placentula</i> Ehrenberg and vars.	893	8.66 (0.16-88.33)	27 (28)	5 (58)	7 (19)	61 (148)
<i>Rhoicosphenia abbreviata</i> (Agardh) Lange-Bertalot	746	4.90 (0.17-59.50)	27 (27)	8 (85)	8 (21)	78 (171)
<i>Navicula cryptotenella</i> Lange-Bertalot ex Krammer et Lange-Bertalot	696	2.12 (0.16-27.06)	28 (25)	9 (115)	5 (17)	55 (129)
<i>Amphora pediculus</i> (Kützing) Grunow	643	3.96 (0.16-50.83)	33 (26)	130 (863)	10 (25)	96 (273)
<b>Rare taxa</b>						
<b>Satellite</b>						
<i>Navicula obsoleta</i> Hustedt	5	0.83 (0.17-1.50)	48 (16)	1 (1)	0 (0)	29 (15)
<i>Rhopalodia gibba</i> var. <i>minuta</i> Krammer	5	0.80 (0.33-1.33)	30 (39)	3 (4)	15 (34)	45 (29)
<i>Eunotia pectinalis</i> (Müller) Rabenhorst	5	0.76 (0.17-1.33)	11 (3)	0 (0)	0 (0)	6 (6)
<i>Stephanodiscus neoastraea</i> Håkansson & Hickel	5	0.70 (0.17-1.33)	59 (41)	65 (98)	49 (45)	517 (415)
<i>Diadesmis laevis</i> (Cleve) Mann	5	0.63 (0.33-0.98)	25 (22)	1 (1)	0 (0)	37 (23)
<b>Urban</b>						
<i>Gomphonema minusculum</i> Krasske	5	10.60 (1.33-43.83)	17 (18)	2 (8)	4 (20)	12 (20)
<i>Caloneis schumanniana</i> var. <i>biconstricta</i> (Grunow) Reichelt	5	2.47 (0.33-5.83)	85 (22)	11 (11)	45 (43)	153 (267)
<i>Stephanodiscus vestibulis</i> Håkansson, Theriot et Stoermer	5	2.46 (0.33-10.48)	40 (35)	103 (58)	29 (30)	370 (174)
<i>Cyclotella bodanica</i> var. <i>lemanica</i> Müller	5	1.63 (0.33-6.67)	34 (11)	1733 (1439)	2 (4)	931 (663)
<i>Synedra filiformis</i> Grunow	5	1.59 (0.83-2.67)	80 (21)	10 (17)	23 (42)	72 (36)
<b>Rural</b>						
<i>Caloneis bacillum</i> (Grunow) Cleve	342	0.99 (0.16-12.00)	34 (28)	7 (21)	10 (24)	66 (166)
<i>Navicula cryptocephala</i> Kützing	293	0.83 (0.16-6.67)	43 (31)	14 (34)	13 (27)	110 (286)
<i>Navicula antonii</i> Lange-Bertalot	264	0.77 (0.17-6.16)	26 (25)	4 (22)	4 (15)	51 (97)
<i>Geissleria acceptata</i> (Hustedt) Lange-Bertalot et Metzeltin	254	0.95 (0.16-9.67)	28 (25)	5 (80)	1 (6)	39 (57)
<i>Mayamaea atomus</i> var. <i>permitis</i> (Hustedt) Lange-Bertalot	229	1.00 (0.16-6.17)	38 (30)	10 (167)	7 (20)	82 (242)

## Discussion

*Rarity metrics* - Contrary to the general expectation for positive correlation between species local abundance and distribution range (Hanski, 1982, 1991; Gaston, 1994), the present dataset did not exhibit such a clear and uniform relationship (Fig. 1. 4). The positive correlation between local abundance and regional distribution is explained most commonly by their interdependence which is considered the rule in nature (Hanski, 1982). There is no generally agreed explanation of this pattern (Hanski and Gyllenberg, 1997), but the two most widely recognized explanations include Brown's niche breadth hypothesis (Brown, 1984) and metapopulation dynamics (Hanski and Gyllenberg, 1993). According to the former explanation, generalist species use ubiquitous resources and are both locally common and widely distributed (core taxa) while specialists use restricted resources and are locally uncommon and have narrow distribution (satellite taxa). According to the latter explanation, locally abundant species become widely distributed because of their low extinction and high colonization rates. The two taxa with the widest distribution ranges (*Achnanthidium minutissimum* (Kützing) Czarnecki (86.22% of sites) and *Cocconeis placentula* Ehrenberg and vars. (90.48% of sites)) also had the highest mean relative abundances (16.86% and 8.66%, respectively). These taxa have cosmopolitan distribution (Patrick and Reimer, 1966; Krammer and Lange-Bertalot, 1991), tolerate a wide range of environmental conditions (Lowe, 1974), they attach firmly to the substrate and are therefore resistant to grazing and scour disturbance (Steinman et al., 1987; Mulholland et al., 1991). In addition, they are early colonizers with high growth rates (Biggs et al.,

1998). On the other hand, satellite taxa are locally uncommon (few of them reached maximum relative abundance above 1%) and have small distribution ranges (1-5 sites). They may be regarded as “true” rare taxa, but their few presences and many more absences obscure any conclusions about their ecology and distribution patterns. According to the “core-satellite species hypothesis” (Hanski, 1982), the majority of taxa should be either core or satellite and with time they can switch between these two stages. However, the expected positive relationship may be due to the fact that it has been derived from small scale studies examining similar organisms with presumed unlimited dispersal abilities, no biotic interactions and few limiting abiotic factors (Brown, 1984). In addition to this predicted pattern, the results from this large-scale study showed a peak in the abundance of urban taxa which makes the relationship look like a ‘U’ shape. It is unclear whether the observed pattern is unique to diatoms or not. Urban taxa are regarded as exceptions from the “core-satellite species hypothesis” with high local growth rates, but with less efficient dispersal (Hanski, 1991). Given the cosmopolitan diatom distribution (Kristiansen, 1996; Finlay et al., 2002), it is hard to believe that dispersal alone is the main reason for the observed pattern. Urban taxa were not nutrient limited, because sites where they occurred did not have significantly lower nutrient concentrations compared to the rest of the sites. Their high relative abundances also exclude the possibility that these taxa may have inferior competitive abilities. Even more, the majority of diatom assemblages with urban taxa were in intermediate or late successional status, implying that they may grow well in such highly competitive and resource limited environments (Stevenson et

al., 1991). Late successional species tend to have high reproductive rates and can accumulate most rapidly when population densities are high (McCormick and Stevenson, 1991). Another possible explanation for the accumulation of urban taxa may be their relatively small size ( $< 50 \mu\text{m}$ ) which may make them less preferred by grazers and more resistant to them (Peterson, 1987).

Satellite taxa may be considered as specialists, if their restricted occurrence and abundance are indication of narrow resource usage. They may have lower growth rates than core taxa (Hanski, 1991) and may become established with difficulty (Söderström, 1989). Theoretically, these taxa may be the most interesting ones to study as indicators of their environments. However, the project's study design was not well suited for characterizing their occurrence and abundance, because not all suitable habitats were examined. Their thorough understanding will require much more intensive sampling and processing methods. For example, the absence of a microbial taxon from a site can be simply due to less efficient sampling methods (Martiny et al., 2006). Reliable identification of a rare diatom is also a challenge. The few observed individuals in a sample may not allow good observation of their morphological features. For example, a species with 1% relative abundance in a sample (the cutoff point between low and high abundance) is only represented with 3 cells in a standard 600 valve count. Also, some taxa with low abundance may have allogenic origin (e. g., washed in from up-stream, present at the site but dead). Less abundant diatoms may be represented by empty cells even months after the species abundance had declined (Owen et al., 1979). In addition, it is very possible that rare diatoms may be

found to have wider distributions after intensive studies focused on them (Finlay et al., 2002; Genkal and Bondarenko, 2006).

Rural species are suspected to exist as transitional stages from core to satellite status or vice versa (Söderström, 1989). Unlike satellite and urban taxa, rural taxa with their wide distribution may be better sampled and represented in the collected composite samples. However, the important environmental variables determining the abundances of rural taxa from Table 1.5 were not uncovered, after an unsuccessful attempt was made to develop regression trees. A close examination of the sites where these taxa occurred revealed that they were not restricted to streams with any specific physical habitat or water quality characteristics. In general, these taxa were found in small mountain streams (1<sup>st</sup>-3<sup>rd</sup> order) at high elevation (>1000 m) with high forest coverage and very low amounts of sand and fines. In addition, all 99 rural taxa occurred in at least two of the three ecoregions with two exceptions. The first one was *Frustulia amphipleuroides* (Grunow) Cleve-Euler which occurred at 31 sites in the Mountain ecoregion with average relative abundances of 0.63% (range 0.17-4.00%). The second exception was *Psammothidium grischunum* fo. *daonensis* (Lange-Bertalot ex Lange-Bertalot et Krammer) Bukhtiyarova et Round which occurred in both Mountain (31 sites) and Xeric (2 sites) ecoregions with abundances of 0.79% (0.17-5.67%) and 0.42% (0.17-0.67%), respectively. The former species is described as acidophilous whereas the latter one is circumneutral, but both taxa are nitrogen-autotrophic, oligosaprobous, oligotraphentic and prefer high oxygen concentrations (Van Dam et al., 1994).

*Rarity predictors* – The only regional rarity metric whose environmental determinants were explored, was the rural metric. Unlike rural taxa, satellite and urban taxa by definition had very low occurrences ( $\leq 5$  sites) and narrow ranges of abundance (0-8 and 0-4 taxa per site, respectively) which made it difficult to model these taxa. Rural taxa richness increased along a longitudinal gradient from mountainous fast-flowing oligotrophic streams with fewer fine substrates to large slow-moving nutrient-rich rivers with abundance of fine substrates (Fig. 1. 5a). Similar “downstream” ecological effect on the distribution of diatom assemblages has been reported from another large scale study, the USGS’s National Water Quality Assessment Program (Potapova and Charles, 2002). The regression tree for the rural metric split the data in two major groups with contrasting environmental conditions (Fig. 1. 5a). The first group corresponded to reference conditions and had low predicted number of rural taxa. Sites in this group had low percentages of sand and fines, low turbidity and phosphorus concentrations, and steep slopes. The majority of sites were mountainous, low order streams in early or intermediate succession with very little total human disturbance in their watersheds (Fig. 1. 6). All these features point to unstable environments with natural disturbance by floods whose frequency, duration, and magnitude shape the local diatom community (Peterson and Stevenson, 1992; Biggs, 1995; Biggs et al., 1999) and reduce species richness (Biggs et al., 1998). Current velocities may create a subsidy-stress gradient (Odum et al., 1979) which affects periphyton community development. Low velocities stimulate algal colonization (Reisen and Spencer, 1970; Peterson and Stevenson, 1989), moderate

current velocities may enhance growth through increased nutrient supply rate (Steinman and McIntire, 1986), while high velocities dislodge periphyton and reset the stages of succession (Peterson et al., 1990; Steinman and McIntire, 1990). In such stressful environments biotic factors are less important in shaping diatom assemblages (Steinman, 1996). The second group on the regression tree corresponded to impacted conditions and had higher predicted number of rare taxa. Sites in this group were characterized with higher percentages of sand and fines, higher turbidity and phosphorus concentrations, and gentle slopes (Fig. 1. 6). The majority of sites were in the Plains and Xeric ecoregions, from a variety of stream orders in intermediate or late succession with various amounts of total human disturbance in their watersheds. All these characteristics are typical of low elevation rivers with stable flows and high nutrient concentrations. Nutrient enrichment may increase algal richness (Porter et al., 2008) and result in higher number of rare taxa (Patrick, 1988). In such stable environments biotic factors may also play an important role in structuring the periphyton assemblage (Steinman, 1996).

*Reference vs. impacted sites* – Rural taxa richness was higher at impacted sites compared to reference ones (Fig. 1. 7a). Most impacted sites were in intermediate and late successional stages (Fig. 1. 6), and had significantly more nutrients compared to reference sites. Late successional stages (Hoagland et al., 1982; Stevenson et al., 1991; Johnson et al., 1997) and high nutrient concentrations (Carr et al., 2005; Stevenson et al., 2008; Veraart et al., 2008) promote algal biomass and diatom richness which provide favorable conditions for the coexistence of multiple species with severe



competition for resources (e. g., nutrients and light). Significant differences in the rural metric at reference and impacted sites in the Mountains may be due to high quality of the reference sites in this region. These conditions may be more representative of their natural state where deviations from it are more likely to be detected (Whittier et al., 2006). In addition, the environmental conditions at reference sites may be better characterized by rural species as indicated by the smaller variation in their number (Fig. 1. 7a). The lack of pronounced difference in the number of rural taxa at reference and impacted sites in the Plains and the Xeric ecoregions may be explained by the fact that they have been under high anthropogenic pressure for much longer periods of time and establishment of true reference sites may not be easily attainable (Whittier et al., 2006).

The number of rural taxa increased with human disturbance. This finding is contrary to other studies which found that the number of rare diatoms decreased along a human disturbance gradient (Potapova and Charles, 2004). The difference in the two studies may be due to the definition of rare taxa. Unlike Potapova and Charles (2004) who defined rare taxa as species which occurred at 10 or less sites and reached at least 2% relative abundance in any site, I used taxa which occurred in 30 or more sites with  $\leq 1\%$  mean relative abundance. Potapova and Charles (2004) explained the lower numbers of rare taxa at higher altitudes with diatom dispersal limitations by mountainous barriers and unstable flow and temperature regime of small streams as opposed to low elevation big rivers, but did not give an explanation for the negative correlation between rare taxa and human disturbance. The positive relationship

between rural taxa richness and human disturbance can be attributed to higher number of ecological niches at impacted mountain sites which had less rural taxa than reference sites in the Plains or the Xeric ecoregions. Such intermediately disturbed environments may provide more resources (Odum et al., 1979) and create additional niches for species coexistence which resulted in significantly higher species richness at these impacted sites.

*Ecological niches* - Contrary to common beliefs that rare taxa inhabit narrow ecological niches and may therefore be sensitive indicators of environmental conditions (Stevenson and Pan, 1999; Magurran, 2004; Potapova and Charles, 2002, 2004), rare taxa in this study did not have narrower ecological niches than common ones. The coefficient of variation for rural taxa resembled closely the one for core taxa, and it was somewhat different from the one for urban and satellite taxa (Fig. 1. 8). The former observation may be due to the larger spatial coverage of both rural and core taxa compared to the limited number of samples where urban and satellite taxa occurred. One possible explanation for the observed overlapping niches for common and rare taxa may be the choice of variables for which they were calculated (e. g., a different set of variables may reveal different patterns). For example, species may differ in terms of substrate preferences, light requirements or flow velocities. However, the rarity of species is unlikely to be a result of one single factor (Gaston, 1994) and there are possibly 5-10 independent niche dimensions for most species (Brown, 1984). In addition, biotic interactions (e. g., competition, grazing) may alter perceptions of niche dimensions. Species niche can be defined as a multidimensional

“hyper-volume” of all environmental variables affecting that species (Hutchinson, 1957), but no single ecological characteristic is totally diagnostic of either common or rare species (Hodgson, 1986). Perhaps there is redundancy in the ecological information provided by common and rare taxa (Marchant, 1999, 2002). As pointed out by Kunin and Gaston (1993), rare and common species have generally broadly overlapping characteristics with some clear differences. For example, causes of commonness and rarity for macroinvertebrates may be due to different biological traits like short life cycles with multiple generations per year for common taxa and low dispersal capabilities for rare taxa (Resh et al., 2005). In his review on the topic, Gaston (1994) concluded that rare species do not necessarily have narrower niches than common ones, but may inhabit more scarce or unusual physical conditions compared to abundant species. The last statement calls for extensive sampling which may never be able to reveal the true species richness in a site (Mao and Colwell, 2005). For this reason, the estimation and detection of rare taxa is not a primary objective in bioassessment programs (Marchant, 2002). The observed overlapping niches for common and rare taxa imply that the assumed better indicator value of rare taxa (Potapova and Charles, 2004) may not be valid. Instead of using rare species autecologies, rare taxa richness may be a better measure of environmental conditions as suggested by the present study.

## **Chapter 2: The importance of live/dead diatoms in bioassessment: a large scale study of Western US streams.**

### **Introduction**

Bioassessment uses different groups of living organisms (fish, macroinvertebrates and algae) as indicators of environmental conditions (Barbour et al., 1999; Carlisle et al., 2008). However, in diatom-based bioassessment, live (containing chloroplast) and dead (empty siliceous cell wall) individuals in an assemblage are not distinguished from each other during the process of conventional diatom analysis. The standard diatom sample processing uses concentrated acid to remove the organic material and to allow diatom identification to the lowest possible taxonomic level using the fine morphological features of their cell walls, which are otherwise obscured by the cell contents. This lack of distinction between live and dead diatoms in the traditional method is resolved by the assumption that the abundance of live diatoms is proportional to the abundance of dead diatoms (e.g. the most abundant live species is also the most abundant dead species), although, dead diatoms might include resident species, euplankton, and dislodged individuals from upstream and impoundments. The lack of distinction between live and dead diatoms may result in overestimated cell density, species diversity and inaccurate characterization of species composition (Pryfogle and Lowe, 1979; Potapova and Charles, 2005; Wilson and Holmes, 1981). However, Stevenson and Bahls (1999) suggested that the percentage of live or dead diatoms in an assemblage may reflect environmental conditions. They

suspected that the percentage of live diatoms may decrease with sedimentation and assemblage age, but this suggestion has not been tested yet.

The proportion of live/dead diatoms in epilithic assemblages is a function of rates of reproduction, death, immigration and emigration (McCormick and Stevenson, 1991; Stevenson and Peterson, 1991), age of the assemblage (Peterson et al., 1990; Steinman and McIntire, 1990), disturbance (Peterson, 1987; Steinman et al., 1987; Peterson and Stevenson, 1992; Peterson, 1996), and resource supply (Steinman and McIntire, 1986; Peterson and Stevenson, 1989; Stevenson et al., 1991). At early stages of development the diatom assemblage is dominated by prostrate and/or erect diatom growth forms (Steinman and McIntire, 1986; Peterson and Stevenson, 1989). With time (Steinman and McIntire, 1990), resources (e. g., light, nutrients) (Patrick, 1967, Steinman and McIntire, 1986; Stevenson et al., 1991) and in the absence of disturbance, the diatom assemblage may develop into a complex three-dimensional structure which resembles terrestrial plant communities (Patrick, 1976; Hoagland et al., 1982; Korte and Blinn, 1983). At later stages of succession the upper layers may prevent the underlying ones from vital resources and increase senescence among epilithic diatoms (Stevenson et al., 1991; Hill, 1996; Johnson et al., 1997), resulting in less live diatoms. In flowing waters succession may be locked in early stages due to frequent disturbance by fast current, substratum instability and grazing activity (Peterson, 1996; Steinman, 1996). Therefore, early- and late-successional species utilize different growth strategies which include high immigration coefficients and high water column abundances for the former and high reproduction rates for the latter

(McCormick and Stevenson, 1991). As the benthic diatom mat grows thicker, it becomes more susceptible to disturbance (Hoagland et al. 1982; Peterson et al., 1990). These dislodged clumps of diatoms become suspended in the water column (drift) and provide continuous supply of potential colonists of newly cleared substrata downstream (Peterson, 1996). The drift levels of live and dead diatoms may fluctuate with diurnal drift patterns, flow and degree of space clearing disturbances (Peterson, 1996; Hamilton and Duthie, 1987). Physical disturbance by storms should not be a concern in this study, because according to protocol, sites were not supposed to be sampled during or shortly after heavy rain events (Herlihy, 2006). The result of all these factors acting upon live diatoms is significant temporal variations in their abundances (Peterson et al., 1990; Steinman and McIntire, 1990).

The structure and function of benthic algal assemblages in streams is determined by small-scale (proximate) factors (e. g., resources, biotic and abiotic stressors) which are constrained by large-scale (ultimate) factors (e. g., climate, geology, land use and biogeography) (Stevenson, 1997). These scale-dependent determinants of benthic algal heterogeneity can be applied to live diatoms in a modified version of Stevenson's model (Fig. 2. 1). The primary proximate resources would be nutrients and light, biotic stressors would include food web structure (piscivore and invertivore fish, and macroinvertebrate grazers) and abiotic stressors would involve stream cross-section as a surrogate for stream size, sediments and watershed disturbance. According to this model, live diatoms in Western US streams and in large homogeneous geographic areas (e. g., ecoregions) will be controlled by a

different set of environmental variables than in hydrologically similar streams (small vs. large) within these areas. I anticipated that resources and macroinvertebrate grazers will stimulate percentage of live diatoms while fish and abiotic stressors will suppress them. The expected effects of the variables from my conceptual model may be confounded by seasonality, successional status and food web structure. Piscivore fish eat invertivore fish which in turn exhibit lower feeding pressure on grazers resulting in their higher abundances (Lamberti, 1996). Grazers, on the other hand, may have positive or negative effects on benthic diatom accrual depending on resource availability (Lamberti, 1996). However, grazer activity will remove senesce material like dead diatoms and clear space for live ones. Therefore, my expectation was that sites with piscivore fish will have more grazers and percentage of live diatoms than sites without piscivore fish.

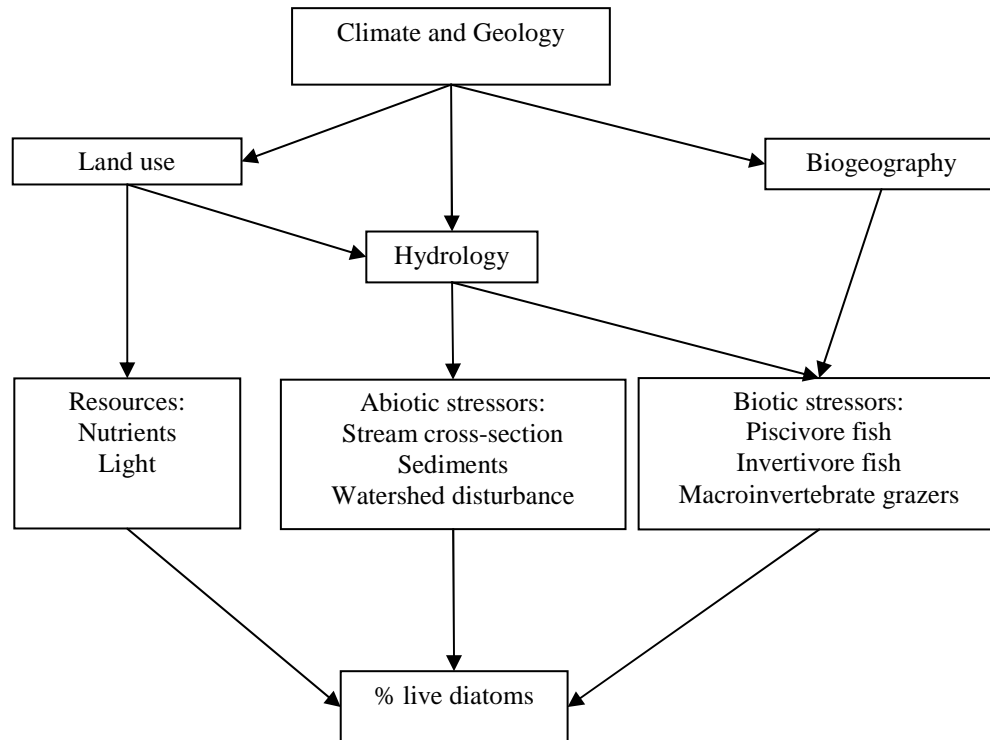


Figure 2.1. Conceptual model for the influence of different environmental factors on the abundance of live diatoms (modified after Stevenson, 1997).

Existing diatom indices use species abundances and ecological preferences (sensitive or tolerant) to infer environmental conditions (Stevenson and Pan, 1999). Many of these indices are good indicators of organic pollution (Lange-Bertalot, 1979), siltation (Bahls, 1993) and eutrophication (Kelly and Whitton, 1995). They are widely used in many regional and national assessment programs which are developing new and improving existing bioassessment tools (Wang et al., 2005; Potapova and Charles, 2007; Weihhoefer and Pan, 2007; Stevenson et al., 2008). These indices require detailed diatom species information obtained from acid-cleaned slides and after considerable taxonomic expertise, but fail to distinguish live/dead diatoms which may



obscure their use as bioindicators. The bioassessment value of live/dead diatoms is unknown. Therefore, this chapter evaluates whether percentage of live/dead diatoms can be used as a metric of human disturbance and explores its properties. Such a metric may not be as detailed as a species level identification, but may provide insight about the “health” of the diatom assemblage which has not been evaluated quantitatively (Lowe and Pan, 1996; Stevenson and Pan, 1999). More specifically, I want to characterize distribution patterns of percentage of live diatoms in Western US streams. In addition, I am also interested in how percentage of live diatoms vary seasonally, with successional development and with food web structure. Next, I want to examine how much variation in the percentage of live diatoms may be explained by the biotic and abiotic factors from my conceptual model. Finally, I want to compare percentage of live diatoms between least and most impacted sites in Western US streams.

## **Materials and Methods**

### *Dataset*

The data used in this study were collected as part of the WEMAP. Sites were selected using a spatially balanced probability-based sampling design (Herlihy et al., 2000). Field sampling followed the WEMAP protocols outlined in the *Background Information*. Environmental conditions (water chemistry, physical habitat and landscape characteristics) and diatom data were collected from perennial streams and rivers in 12 Western states of the US. The sites in this study were sampled once during

the summer low flow conditions (May-October) from 2000 to 2002. A total of 587 benthic samples with a full set of biological data and non-zero live and dead diatom abundances were analyzed (Fig. 2.2).

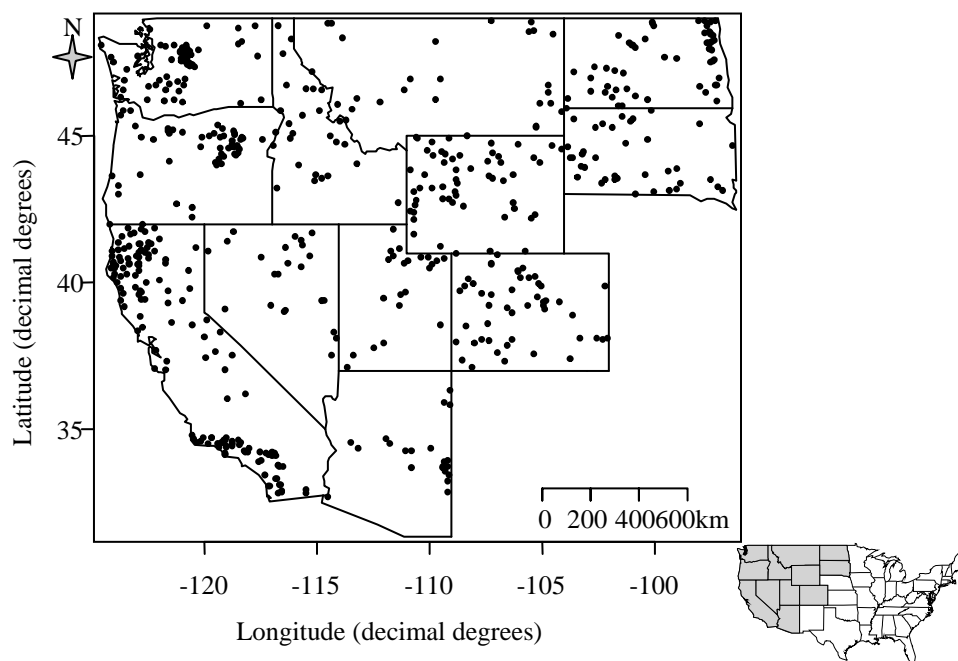


Figure 2.2. Maps of USA and sampling sites (n=587) in 12 states in the Western USA.

### *Lab analysis*

Each composite algal sample (collected as in Chapter 1) was split in two subsamples. A small amount of the first subsample was pipetted into a Palmer-Maloney counting chamber and examined at 400x under a light microscope. A minimum of 300 natural algal counting units (NCU, natural grouping of algae, i.e., each individual filament, colony, or isolated cell) were counted and identified (Acker, 2002). Diatoms were counted as either live (possessing visible cell contents), in which

case their number contributed to the total of 300 NCUs, or dead (empty cells) and their number was also recorded but not included in the total of 300 NCUs. When a sample was dominated by diatoms, all 300 NCUs could be live diatoms. When a sample was too sparse, fewer than 300 NCUs were recorded. Part of the second algal subsample was acid-digested and used for the preparation of one permanent diatom slide (Patrick and Reimer, 1966). A minimum of 600 diatom valves were counted and identified to the lowest possible taxonomic level (mainly species) at 1000x under a Leica DM LB light microscope. The diatom taxonomy followed primarily Krammer and Lange-Bertalot (1986, 1988, 1991a, b) and Patrick and Reimer (1966, 1975).

#### *Data analysis*

Percentage of live diatoms (% live diatoms) was calculated as (number of live diatom cells/(number of live diatom cells + number of dead diatom cells))\*100. The percentage live diatoms were selected over the ratio of live/dead diatoms, because it has been suggested that percentages are less variable than ratios (Fore et al., 1996). For example, the same live/dead ratio may result from a pair of very large or very small values, while % live diatoms may represent processes at the assemblage level.

*Distribution patterns of live diatoms* – To characterize the distribution of % live diatoms not only across the Western US, but also at different spatial scales, the data were summarized for each one of three aggregated level III ecoregions (MT-Mountains, PL-Plains, XE-Xeric) (Omernik, 1987; US EPA, 1996) and two stream order classes (1<sup>st</sup>-3<sup>rd</sup> order and > 3rd order; Strahler, 1957) within each ecoregion.

*Seasonality* - An important factor that may affect the data interpretation is seasonality in assemblage development. To evaluate whether sampling time of the year resulted in different assemblages, the species composition and % live diatoms between two within-year site visits were compared for 27 sites. Second time visits were not part of the analyzed dataset. The species composition from the traditional diatom analysis for the two within-year site visits was plotted on non-metric multidimensional scaling plot, using the Bray- Curtis similarity coefficient as a distance measure (Bray and Curtis, 1957). This coefficient varies between 0-100% (0 – no similarity, 100 – total similarity). If within-year site visits have high similarity coefficients, their distances on the ordination plot would be small. Percentage live diatoms between 1<sup>st</sup> and 2<sup>nd</sup> time visits were compared with a paired t-test.

*Successional stage* - To assess whether % live diatoms may be influenced by assemblage development, sites were classified as being in early and late successional stages based on the sum of relative abundances of species with ruderal (R) strategy (Biggs et al., 1998). These species are early colonizers of denuded habitats and reach high abundances at the start of succession. Sites with more than 50% relative abundance of R strategy species were classified in early succession, while sites with less than 10% relative abundance of R strategy species were classified in late succession. This classification was based on the species composition from the traditional diatom analysis, because live diatoms were not identified to species. Percentage live diatoms at early and late successional sites were compared with a t-test. When the assumptions of the t-test were not met (e. g., normality, equal variance),

the data were transformed, and if that did not help meet the test assumptions, a non-parametric test was used (e. g., Mann-Whitney test). I anticipated that late successional stages will have less live diatoms due to resource limitations with assemblage development (Stevenson et al., 1991; Johnson et al., 1997).

*Food web structure* - The effect of food web structure on % live diatoms was evaluated by looking at piscivores presence/absence and at piscivore density (high vs. low). My initial expectation was that sites with piscivores or with high piscivore density will have lower invertivore density than sites without piscivores or with low piscivore density. Lower invertivore density would increase grazers density and % live diatoms. To test this hypothesis, all sites with fish (437) were split in two groups, with (60) and without (375) piscivores. Second, sites with piscivores (60) were divided into 3 groups of high, medium and low density (18, 22 and 20 sites, respectively). Next, % live diatoms at sites with and without piscivores, and at sites with high and low piscivore density were compared with a t-test. The same comparison was done for chlorophyll *a* concentrations, invertivore fish density and grazers density. All data were log-transformed.

*Conceptual model* - I used multiple regression to evaluate my conceptual model. Multiple regression models the relationship between the response variable (% live diatoms) and the predictor variables (environment) to find out which predictor variables best describe the pattern in the distribution of the response variable. A good predictive model has a high adjusted  $R^2$  (a measure of model's goodness-of-fit). Adjusted  $R^2$  accounts for the number of variables in the model, because it penalizes

models with higher number of variables. However, in this study multiple regression was used for hypothesis testing and the significance of the regression coefficients (variables' slopes) was considered more important than model's adjusted  $R^2$ . Multiple regression models assume linear relationship between the response and predictor variables, normally-distributed and independent errors with constant variances, and no outliers or influential observations that can change the model fit. All predictor variables from my conceptual model (Table 2.3) were scaled to make them comparable, while the response variable was square-root transformed to normalize it. To evaluate whether the data met the model assumptions, I did diagnostic check of the residuals by examining residuals vs. fitted values plots, normal QQ-plots and leverage plots. Influential points that fell outside of the 0.5 Cook's distance envelope, were removed (two sites in XE (1<sup>st</sup>-3<sup>rd</sup> order) and two sites in XE (> 3<sup>rd</sup> order)). Multiple regression can not handle missing values and after deleting them from the dataset, it was reduced to 459 sites.

*Reference vs. impacted sites* – To evaluate whether % live diatoms can be used as a metric of human disturbance and be able to distinguish between least- (reference) and most-impacted (impacted) sites, they were compared with a t-test. This comparison was done for each of the three ecoregions (MT, PL and XE) and two stream order classes (Strahler, 1957; 1<sup>st</sup>-3<sup>rd</sup> order and > 3<sup>rd</sup> order) within each ecoregion. EPA provided reference designation for each site. The sets of criteria used to classify sites are described in the *Materials and Methods* section of Chapter 1. My initial expectation was that % live diatoms would decrease with human disturbance.

All data analyses were performed with R language and environment for statistical computing (R Development Core Team, 2008).

## **Results**

Environmental conditions varied substantially among the sampled streams (Table 2.1). In general, streams in the MT ecoregion were small, steep, with fast flow, located at high elevation and with high annual precipitation. They were forested and well-shaded, with minimum human disturbance in their watersheds. Mountain streams had low nutrients (total nitrogen and total phosphorus), fines and turbidity levels. In contrast, streams in the PL ecoregion were on average larger, with gentle slopes and slow current, and were located at low elevation with low annual precipitation. Plains' streams were less forested and shaded, and had high degree of human disturbance in their watersheds. In addition, streams in the PL had high nutrients, fines and turbidity levels. Compared to the MT and PL, streams in the XE ecoregion fell somewhere between them, but with a few exceptions for some variables. For example, XE streams had on average the largest watershed area, the highest total nitrogen concentrations, and the highest human population density. The lowest mean value of chlorophyll *a* was recorded from MT streams ( $17.96 \text{ mg/m}^2$ ), followed by PL and XE ( $34.49$  and  $44.26 \text{ mg/m}^2$ , respectively). Small ( $1^{\text{st}}$ - $3^{\text{rd}}$  order) and large ( $> 3^{\text{rd}}$  order) streams in all three ecoregions also differed in terms of the described variables. For the most part small streams had some of the characteristics of MT sites, while large streams resembled PL sites. Exceptions to this trend were some variables for the Xeric

ecoregion which were higher for small streams than for large ones (e. g., watershed area, population density, nutrients, chlorophyll *a* concentrations).



Table 2. 1. Environmental variables description and summaries for the whole dataset (West), three ecoregions (MT-Mountains, PL-Plains, XE-Xeric), and two stream order classes (1st-3rd and > 3rd).

Variable code	Variable (units)	West (n=587)	MT (n=356)	PL (n=124)	XE (n=107)
<b>Algal biomass</b>					
CHL_M2	Chlorophyll <i>a</i> (mg/m <sup>2</sup> )	26 (0-327)	18 (1-273)	34 (0-327)	44 (2-277)
<b>Water quality</b>					
COND	Conductivity (uS)	656 (9-12290)	203 (9-2757)	1710 (120-6258)	942 (23-12290)
NTL	Total nitrogen (ug/L)	1012 (15-89750)	182 (15-3720)	1494 (114-18775)	3214 (60-89750)
PHSTVL	pH	8 (6.2-9.9)	7.8 (6.2-9.1)	8.3 (7.3-9.9)	8.2 (6.9-9.3)
PTL	Total phosphorus (ug/L)	82 (0-2549)	24 (0-568)	217 (2-2549)	117 (2-2165)
SIO2	Silica (mg/L SiO <sub>2</sub> )	18 (0-78)	17 (1-53)	16 (0-57)	25 (4-78)
TURB	Turbidity (NTU)	7 (0-399)	1 (0-64)	28 (1-399)	4 (0-46)
<b>Physical habitat</b>					
FLOW_CFS	Instantaneous discharge (ft <sup>3</sup> /sec)	13 (0-929)	15 (0-929)	11 (0-143)	10 (0-404)
PCT_FN	Substrate fines -- silt/clay/muck (%)	22 (0-100)	11 (0-100)	49 (0-100)	33 (0-100)
XCDENBK	Bank canopy density - mean (%)	70 (0-100)	77 (0-100)	56 (0-98)	63 (0-100)
XDEPTH	Thalweg depth - mean (cm)	33 (2-159)	30 (2-155)	48 (4-159)	29 (3-98)
XWIDTH	Wetted width - mean (m)	5.7 (0.3-42.5)	5.1 (0.4-42.3)	7.8 (0.5-42.5)	5.3 (0.3-35.4)
XWXD	Mean width/depth ratio (m/m)	2.7 (0-34.1)	2.3 (0-34.1)	4.5 (0-25.1)	2.2 (0-25.1)
<b>Landscape characteristics</b>					
AG_TOT	Agriculture in watershed (%)	11 (0-98)	1 (0-66)	48 (0-98)	7 (0-98)
FOR_TOT	Forest in watershed (%)	49 (0-100)	69 (0-100)	8 (0-89)	33 (0-95)
POPDENKM	Population density (persons/km <sup>2</sup> )	9 (0-724)	2 (0-97)	6 (0-358)	37 (0-724)
PRECIP_M	Annual precipitation (m)	0.9 (0.1-3.7)	1.1 (0.4-3.7)	0.4 (0.3-0.6)	0.5 (0.1-1.4)
SLOPMAN	Watershed slope - mean (%)	13 (0-36)	17 (3-36)	3 (0-15)	13 (1-29)
TMEAN_YR	Air temperature (mean °C/year)	7 (-1-23)	7 (-1-18)	6 (2-12)	11 (3-23)
WSAREA	Watershed area (km <sup>2</sup> )	3161.2 (0.1-636285.2)	113.7 (0.2-9460.5)	3702.8 (8.5-52720.8)	12672.8 (0.1-636285.2)
XELEV	Elevation at site (m)	1197 (36-3660)	1357 (10-3660)	726 (240-2149)	1210 (36-2627)

NTU-nephelometric turbidity units

*Distribution patterns of live diatoms* - The average % live diatoms in a sample was 34.50 % (range 2.08-97.02 %) (Table 2.2). The highest mean % live diatoms were in the MT ecoregion (36.38%) and they decreased from the XE (35.49%) to the PL ecoregion (28.27%). In all three ecoregions, mean % live diatoms and their ranges were higher in small order streams (1<sup>st</sup>-3<sup>rd</sup>) than in larger ones (>3<sup>rd</sup>). Reference and impacted sites had an average of 33.65% (range 8.56-90.84%) and 34.87% (range 2.08-97.02%) live diatoms, respectively.

Table 2.2. Summary for sampled sites (n) and % live diatoms in the dataset (West), three ecoregions (MT-Mountains, PL-Plains, XE-Xeric) and two stream order classes (1st-3rd and > 3rd).

<b>Region</b>	<b>n</b>	<b>Mean</b>	<b>Min</b>	<b>Max</b>
West	587	34.50	2.08	97.02
MT	356	36.38	2.44	92.48
PL	124	28.27	2.08	86.68
XE	107	35.49	4.91	97.02
MT (1st-3rd)	324	36.45	2.44	92.48
MT (> 3rd)	32	35.64	11.76	66.96
PL (1st-3rd)	50	30.85	3.45	86.68
PL (> 3rd)	74	26.53	2.08	76.84
XE (1st-3rd)	84	35.95	4.91	97.02
XE (> 3rd)	23	33.84	12.98	68.00
Reference	130	33.65	8.56	90.84
Impacted	135	34.87	2.08	97.02

*Seasonality* – Seasonality may have influenced both species composition and % live diatoms. The average Bray-Curtis similarity between two site visits was 56.00% (range 26.20-80.60%). The two-dimensional ordination plot did not represent the true site differences very well (stress value = 26.51; Fig. 2.3) and its interpretation was treated with caution. Despite this relatively large variation in species similarity between two site visits, there was no significant difference in % live diatoms between visits ( $p$ -value=0.81). The mean difference in % live diatoms between visit 2 and visit 1 was 1.04 (range -59.36-53.03%). However, there was a large amount of variability in % live diatoms with time (Fig. 2.4).

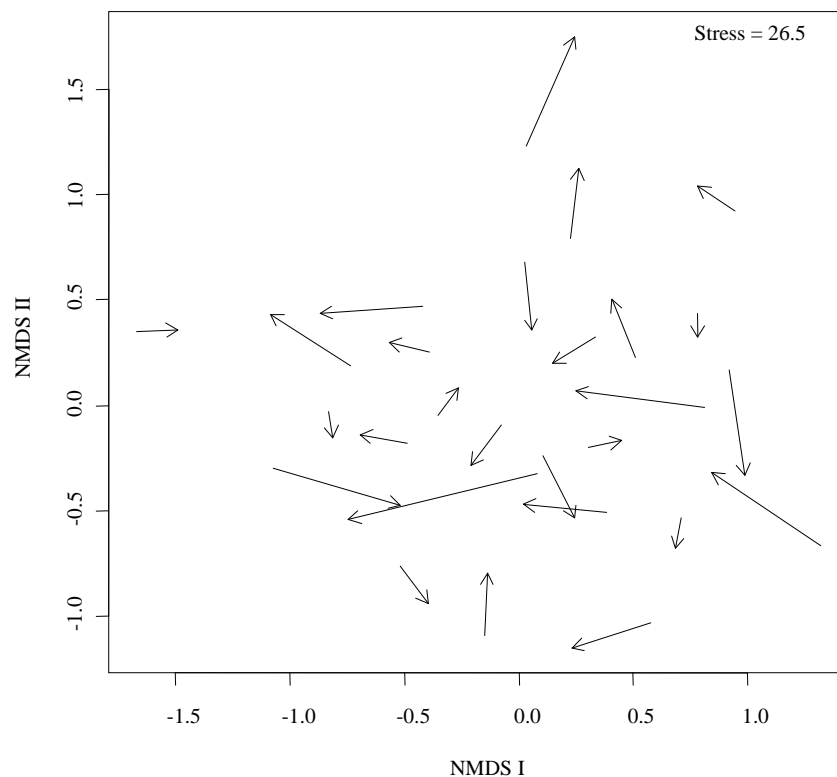


Figure 2.3. Non-metric multidimensional scaling plot for diatom species composition between two within-year site visits ( $n=27$ ). The arrow length corresponds to Bray-Curtis similarity distance. Arrows point in the direction of the second site visit.

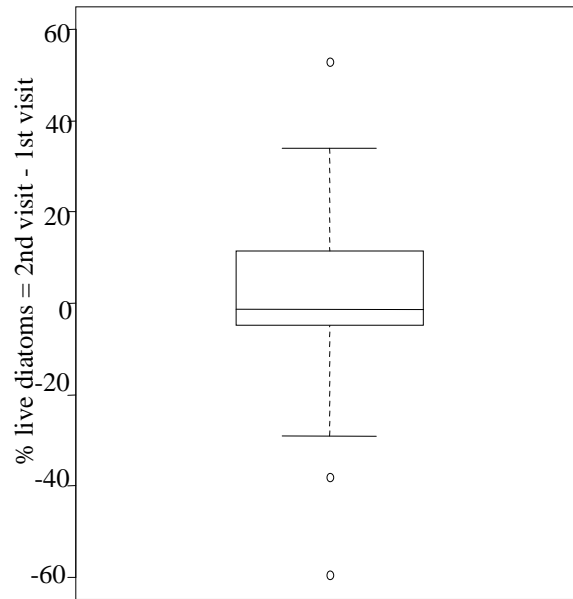


Figure 2.4. Boxplot for the difference in % live diatoms between 2<sup>nd</sup> and 1<sup>st</sup> within-year site visits (n=27).

*Successional stage* - There was no significant difference (p-value=0.43) in the % live diatoms between early and late successional sites (Fig. 2.5). Overall, % live diatoms tended to increase at late successional sites, but they were significantly higher than early successional sites only in the MT ecoregion (p-value=0.02). Late successional sites in the Mountains were characterized with significantly higher amounts of total phosphorus, substrate fines, turbidity and % agriculture in their watersheds (Fig. 2.6). On the contrary, early successional sites were steeper, better shaded and less impacted by human activities. When MT sites were further split into small (1<sup>st</sup>-3<sup>rd</sup> order) and large (>3<sup>rd</sup> order) streams, the significant difference in % live

diatoms between early and late successional sites was still evident, but only for small streams (Fig. 2.7). Early and late successional sites in these small streams had the same differences in their environmental variables as in the MT ecoregion (Fig. 2.8). Small and large streams within the other two ecoregions (PL and XE) did not exhibit significant differences in the % live diatoms (Fig. 2.7).

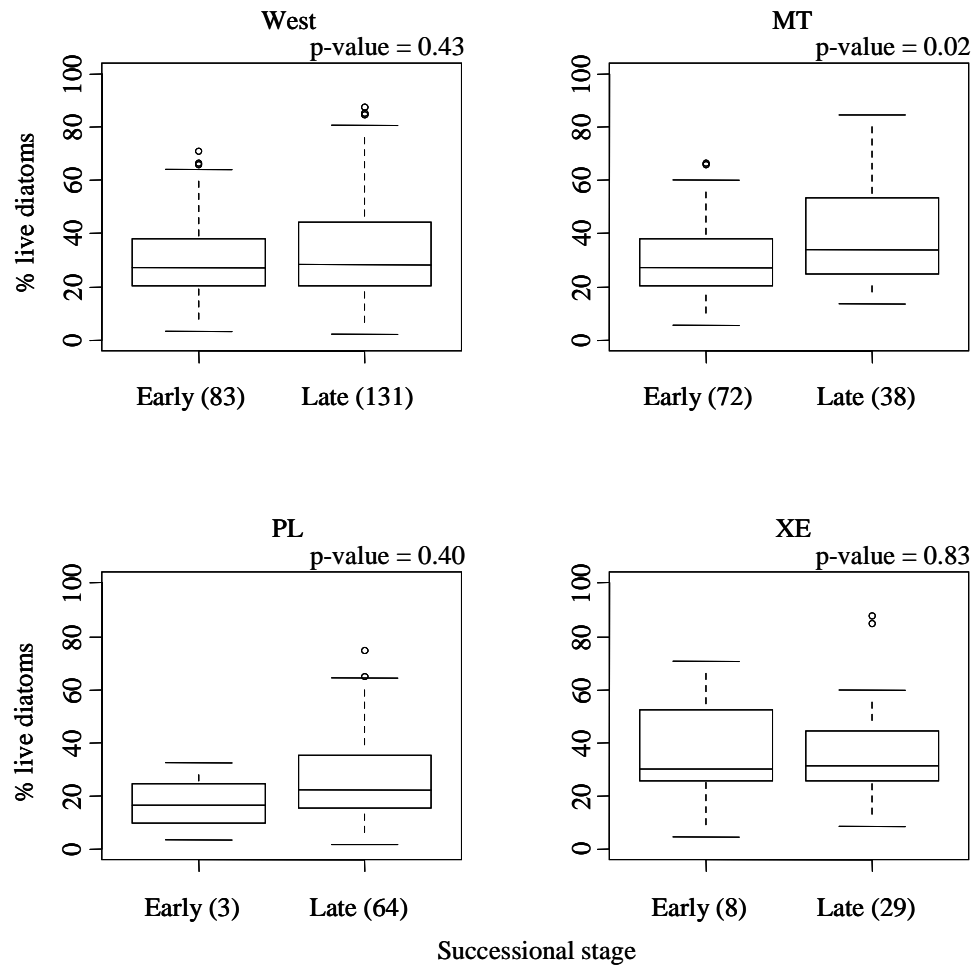


Figure 2.5. Boxplots for early and late diatom successional sites (numbers in parentheses) in the dataset (West) and three ecoregions (MT-Mountains, PL-Plains and XE-Xeric) with p-values from t-tests.

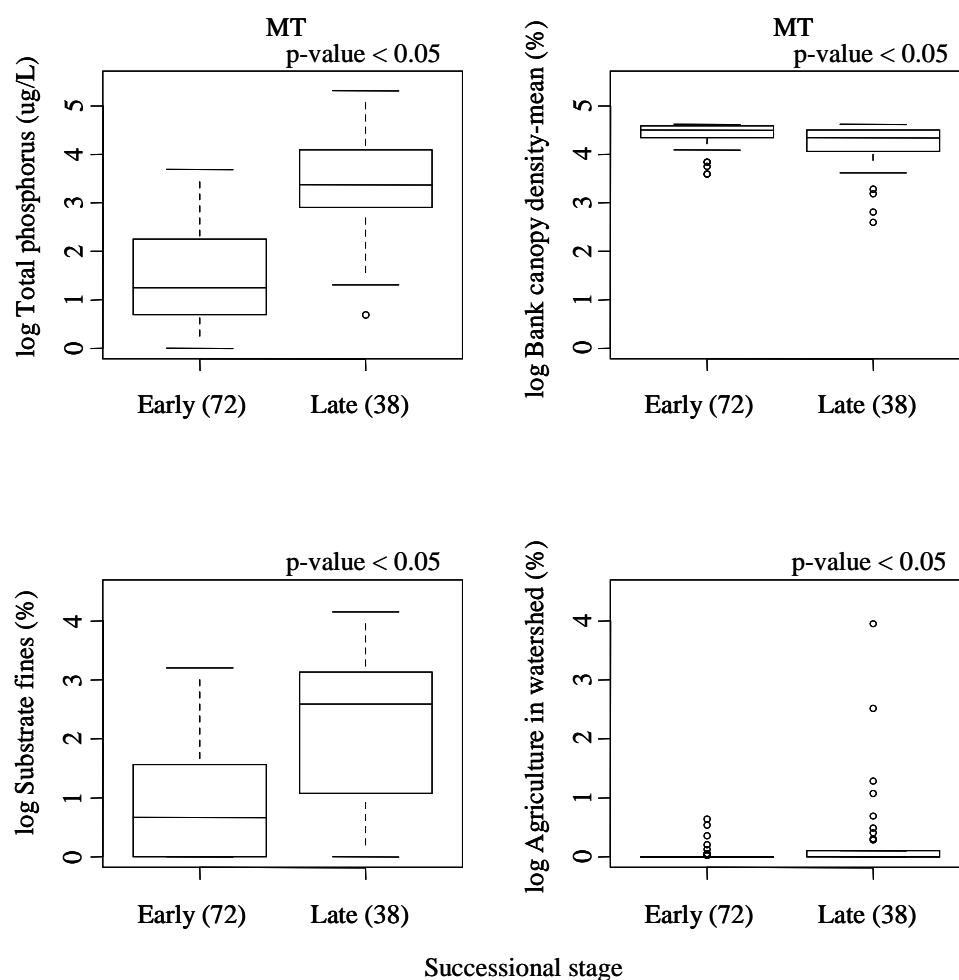


Figure 2.6. Boxplots for selected environmental variables from the conceptual model for early and late diatom successional sites (numbers in parentheses) in the MT ecoregion.

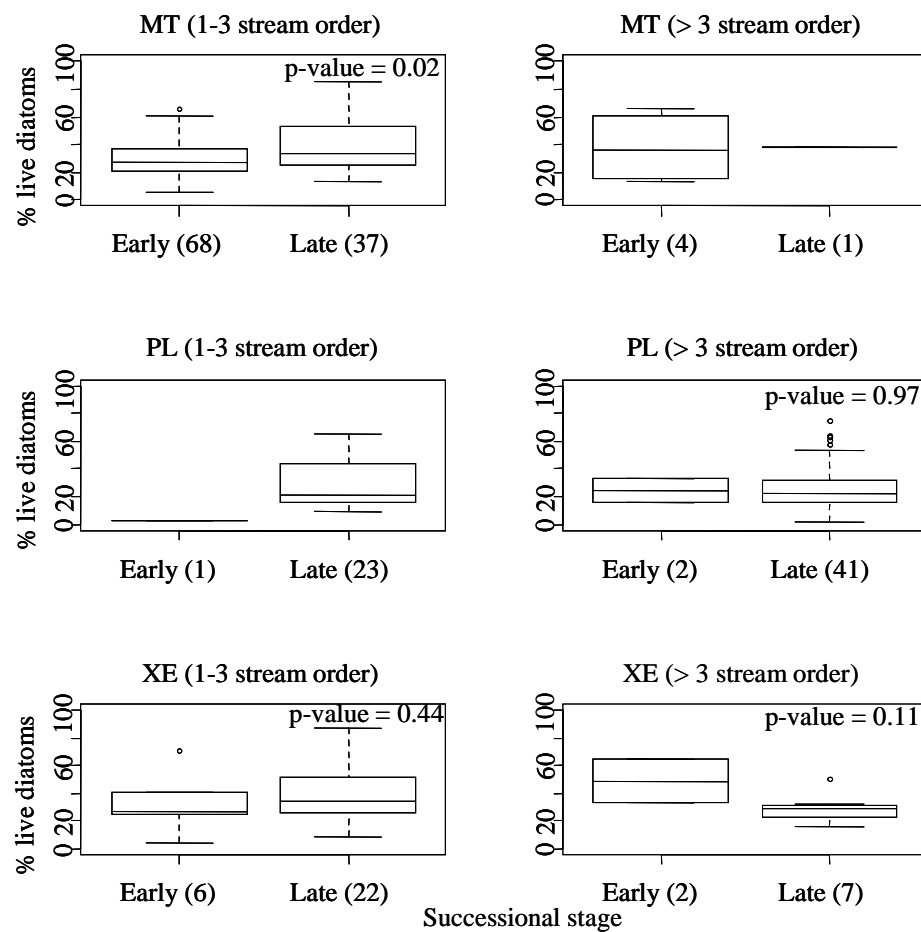


Figure 2.7. Boxplots for early and late diatom successional sites (numbers in parentheses) at low (1<sup>st</sup>-3<sup>rd</sup>) and high (>3<sup>rd</sup>) order streams in three ecoregions (MT-Mountains, PL-Plains and XE-Xeric) with p-values from t-tests.

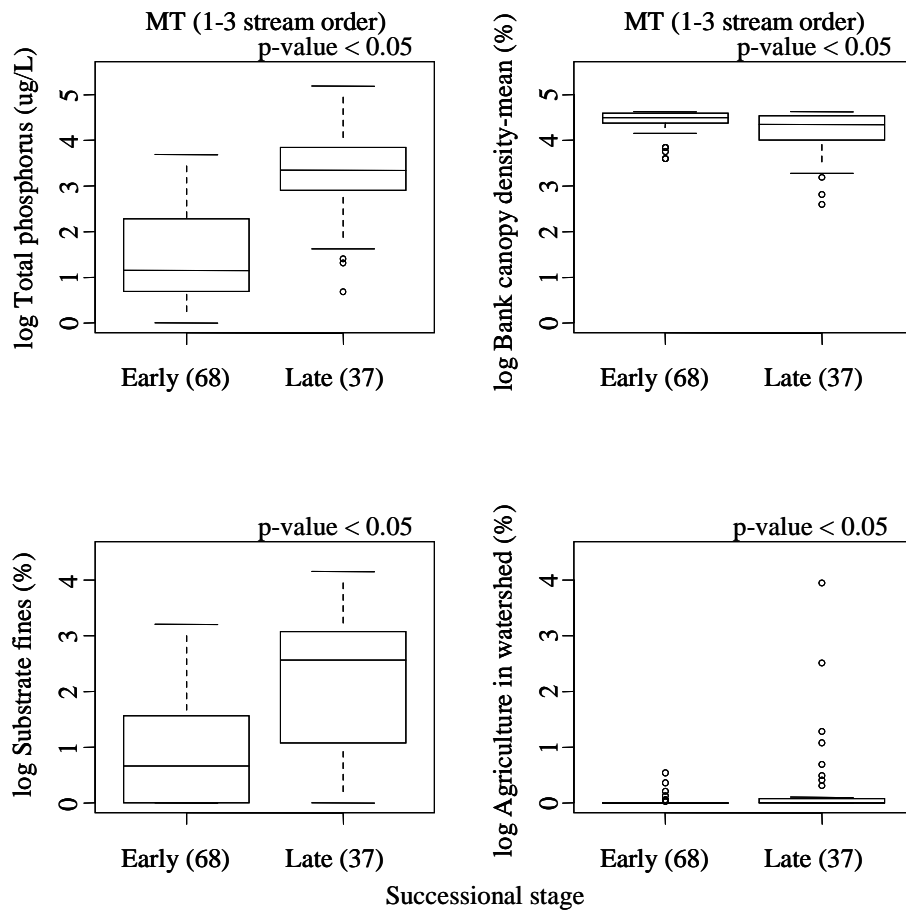


Figure 2.8. Boxplots for selected environmental variables from the conceptual model for early and late diatom successional sites (numbers in parentheses) in small (1<sup>st</sup>-3<sup>rd</sup> order) streams in the MT ecoregion.

*Food web structure* – The mean number of fish individuals per site was 171 (range 1-2549). Sites with piscivores had on average 293 (range 1-2549) fish individuals compared to sites without piscivores which had 152 (range 1-2139) fish individuals. Piscivore density was very low with a mean of 0.005 (range 0-0.045) individuals/m<sup>2</sup>, and was even lower for low (mean 0.0003, range 0-0.0007) and high (mean 0.01, range 0.003-0.045) piscivore density sites. However, low piscivore density sites had on average 462 (range 1-2549) fish individuals compared to high



piscivore density sites which had 71 (range 12-180) fish individuals. The presence of piscivore fish resulted in significantly lower ( $p\text{-value} < 0.05$ ) grazers density (Fig. 2.9) and % live diatoms (Fig. 2.10). There was no significant difference in invertivore fish density or chlorophyll *a* concentrations at sites with and without piscivores. Sites with piscivores were larger streams with more nutrients (total nitrogen and total phosphorus), fine substrates, higher turbidity, more agriculture and bigger watershed areas. Streams without piscivores were steeper and at higher elevation, well-forested and well-shaded. Sites with high piscivore density had significantly higher grazers density ( $p\text{-value} = 0.01$ ) compared to sites with low piscivore density. Sites with low piscivore density were larger, at higher elevation, with faster stream flow and larger watershed areas.

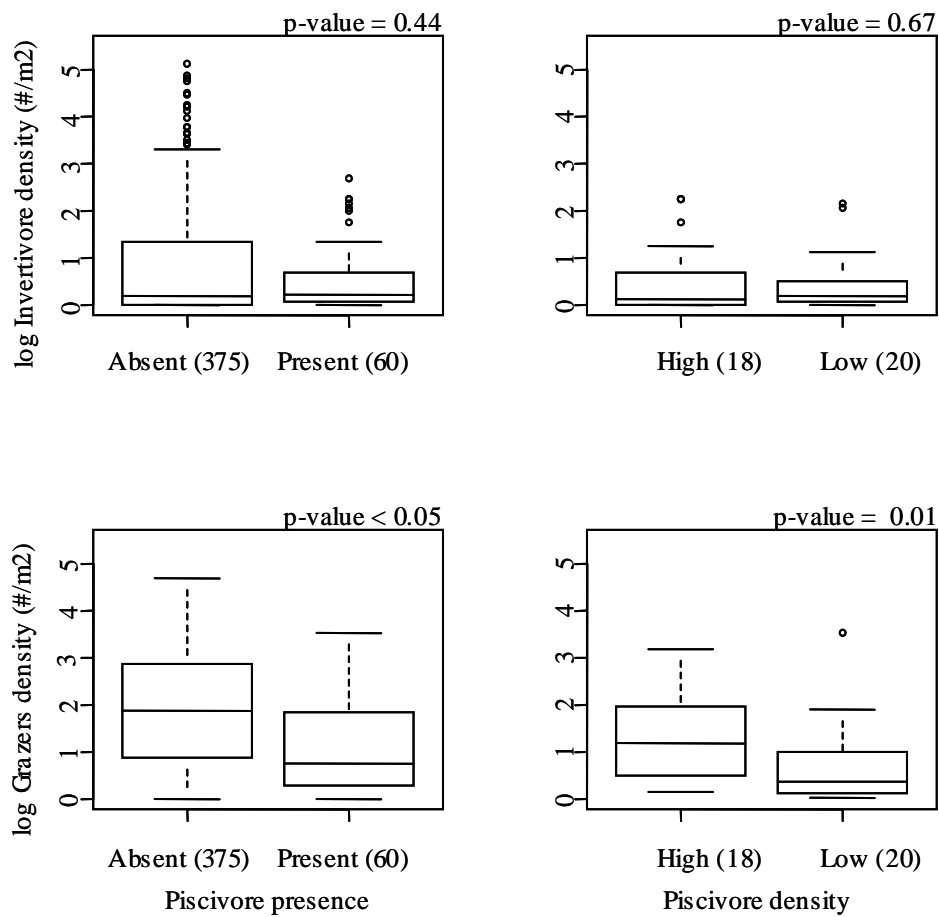


Figure 2.9. Boxplots for invertivore fish and macroinvertebrate grazers density at sites (numbers in parentheses) with and without piscivore fish with p-values from t-tests.

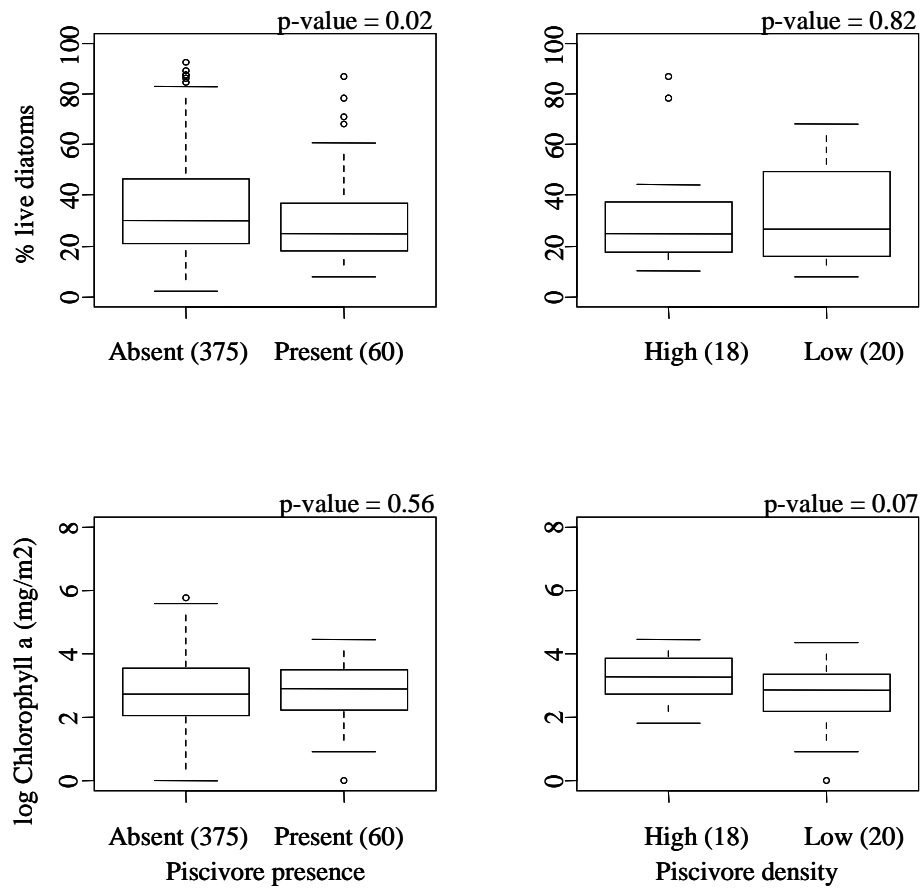


Figure 2.10. Boxplots for % live diatoms and chlorophyll *a* concentrations at sites (numbers in parentheses) with and without piscivore fish with p-values from t-tests.

*Conceptual model* - Variability of % live diatoms in Western US streams was poorly associated with biotic and abiotic factors from my conceptual model (adjusted  $R^2 = 0.05$ , Table 2.3). Stream size and % agriculture in the watershed were negatively related to % live diatoms. Model performance was not much improved at finer spatial scales (e. g., ecoregions). The explained variance for % live diatoms in the MT and XE ecoregions was only 6% and 4%, respectively. No model was developed for the PL ecoregion. As expected, different sets of environmental variables were important

determinants of % live diatoms at different spatial scales. In addition to stream cross-section and % agriculture, piscivore fish density determined % live diatoms in the MT. Percentage live diatoms in the XE ecoregion were inversely related to bank canopy density and substrate fines. When MT and XE ecoregions were further split into small (1st-3rd order) and large streams ( $> 3^{\text{rd}}$ ), the explained variance in % live diatoms increased. Percentage live diatoms in small MT streams were positively affected by substrate fines and piscivore density, but negatively influenced by stream size and % agriculture. On the contrary, % live diatoms in large MT streams were inversely related to substrate fines. The regression models for the two stream classes in the XE ecoregion had the highest predictive power (Table 2.3). Total phosphorus, bank canopy density, fine substrates and stream size explained 21% of the variation in % live diatoms for small xeric streams. Total phosphorus, stream size, grazers and piscivore density were positively related to % live diatoms in large xeric streams and explained 55% of their variation.

Table 2.3. Results from multiple regression for the whole dataset (West), two ecoregions (MT-mountains and XE-xeric) and two stream order classes (1st-3rd and > 3rd). No model was developed for the PL (Plains). “+” and “-” signs represent significant positive or negative relationship, respectively. Numbers in parentheses denote number of sites.

Conceptual model variables	Variables (units)	West (459)	MT (257)	MT (1st-3rd) (227)	MT (> 3rd) (28)	XE (83)	XE (1st-3rd) (61)	XE (> 3rd) (18)
Nutrients	Total phosphorus (ug/L)						-	+
Shade	Bank canopy density - mean (%)					-	-	
Sediments	Substrate fines –silt/clay/muck (%)			+	-	-	-	
Stream cross-section	Mean width*depth product (m <sup>2</sup> )	-	-	-			+	+
Watershed disturbance	Agriculture in watershed (%)	-	+	-				
Macroinvertebrates	Grazers density (#/m <sup>2</sup> )							+
Invertivore fish	Invertivore fish density (#/m <sup>2</sup> )							
Piscivore fish	Piscivore fish density (#/m <sup>2</sup> )		+	+	RM			+
	Adjusted R <sup>2</sup>	0.05	0.06	0.09	0.10	0.04	0.21	0.55

RM - variable removed, because of colinearity with total phosphorus

*Reference vs. impacted sites* - There was no significant difference in % live diatoms between reference and impacted sites in the studied Western US streams (Fig. 2.11). The difference in % live diatoms between reference and impacted sites was significant only in the MT (p-value = 0.02) and PL (p-value = 0.05) ecoregions, and in small MT streams (1<sup>st</sup>-3<sup>rd</sup> order, p-value = 0.02, Fig. 2.13). However, % live diatoms exhibited opposite patterns in the two ecoregions. There were more live diatoms at impacted sites in the MT and less in the PL, despite their environmental similarities. For example, reference sites in the MT and PL were well-forested and shaded while impacted sites in both ecoregions had more nutrients (total nitrogen and total phosphorus), fine substrates and higher turbidity levels (Fig. 2.12). Reference and impacted sites in small MT streams (1<sup>st</sup>-3<sup>rd</sup> order) also exhibited these similar environmental differences (Fig. 2.14).

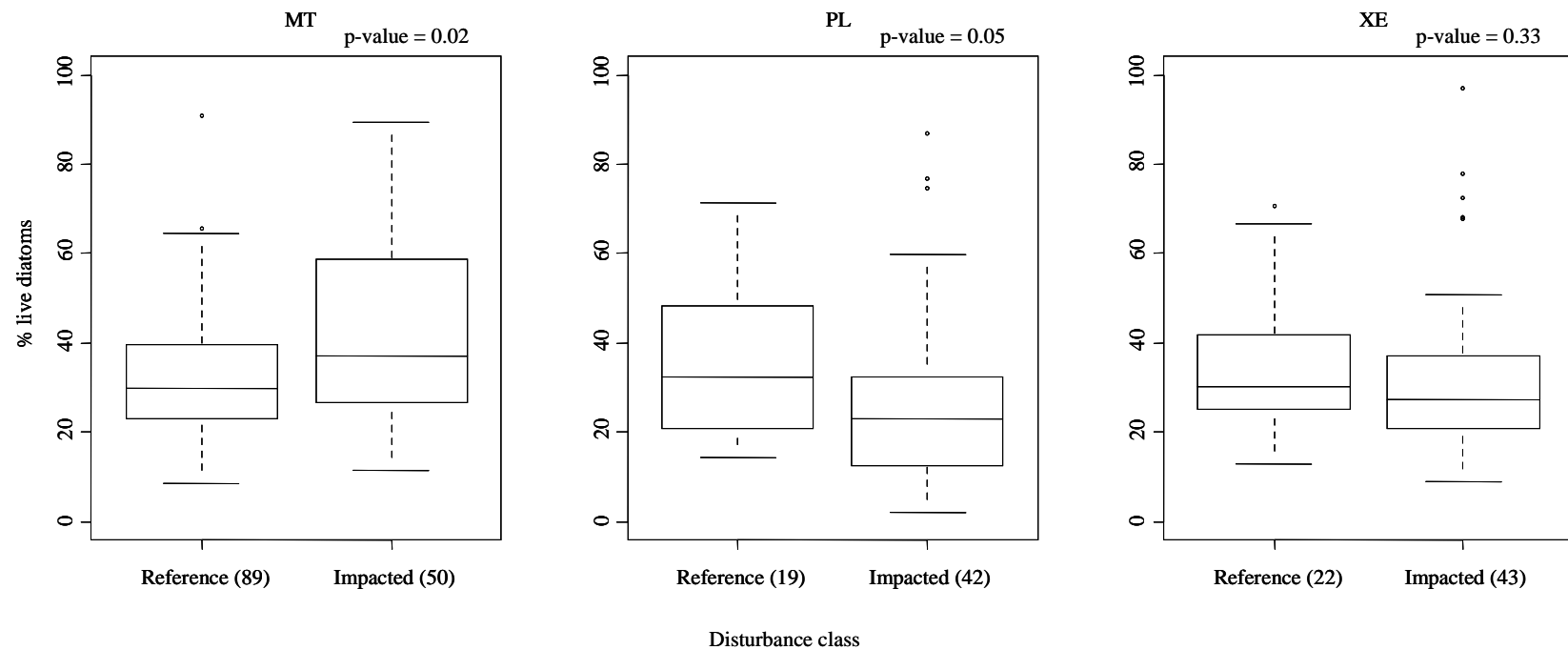


Figure 2.11. Boxplots for reference and impacted sites (numbers in parentheses) in three ecoregions (MT-Mountains, PL-Plains and XE-Xeric) with p-values from t-tests.

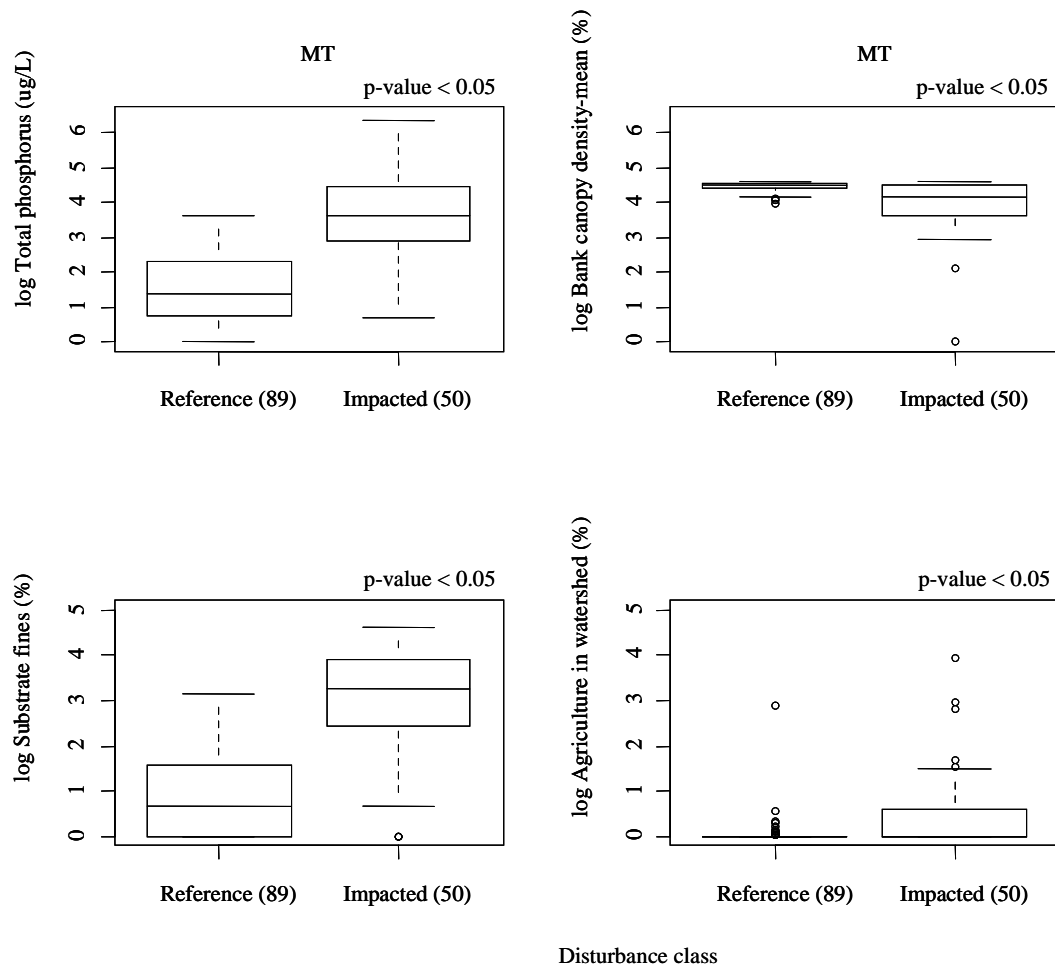


Figure 2.12. Boxplots for selected environmental variables from the conceptual model for reference and impacted sites (numbers in parentheses) in the MT ecoregion.



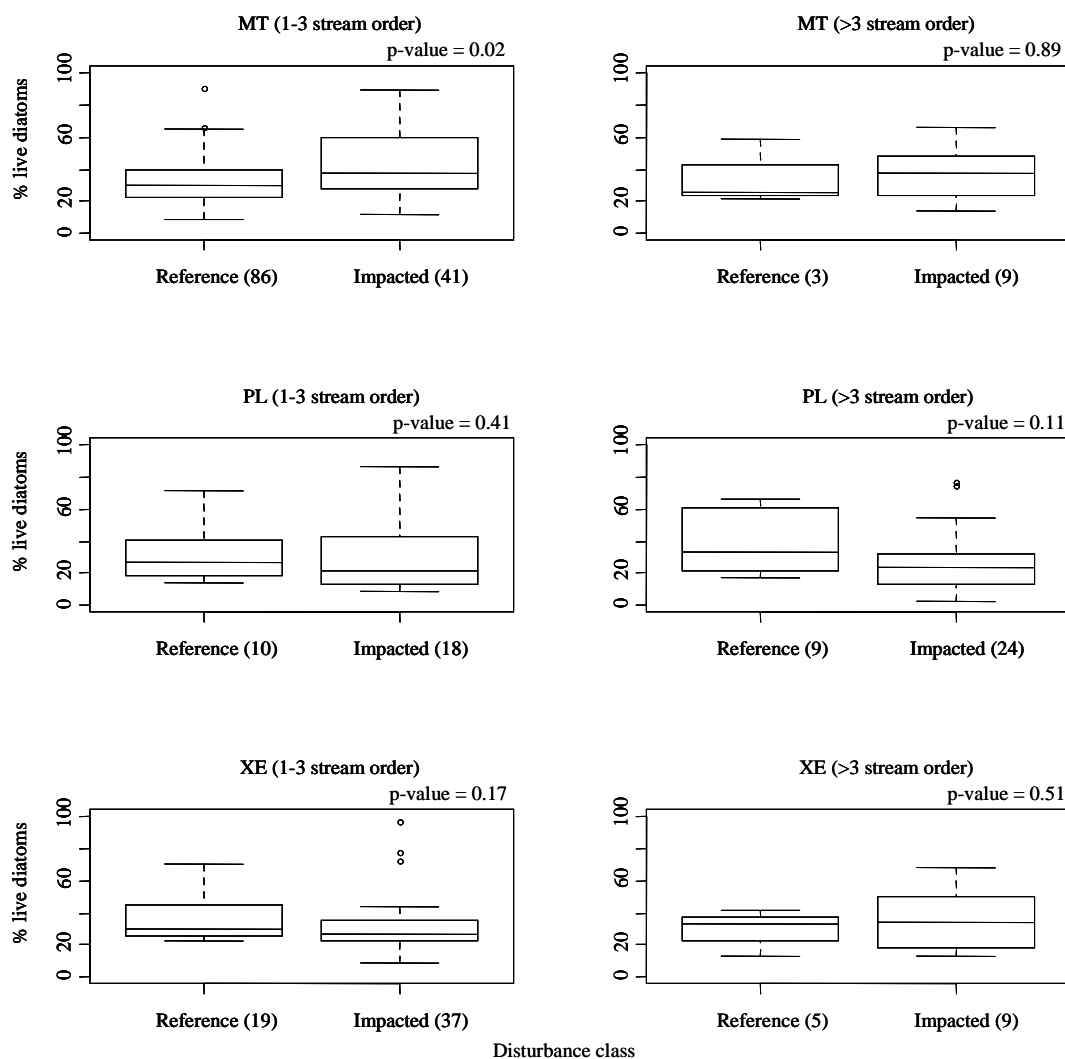


Figure 2.13. Boxplots for reference and impacted sites (numbers in parentheses) at low (1<sup>st</sup>-3<sup>rd</sup>) and high (>3<sup>rd</sup>) order streams in three ecoregions (MT-Mountains, PL-Plains and XE-Xeric) with p-values from t-tests.

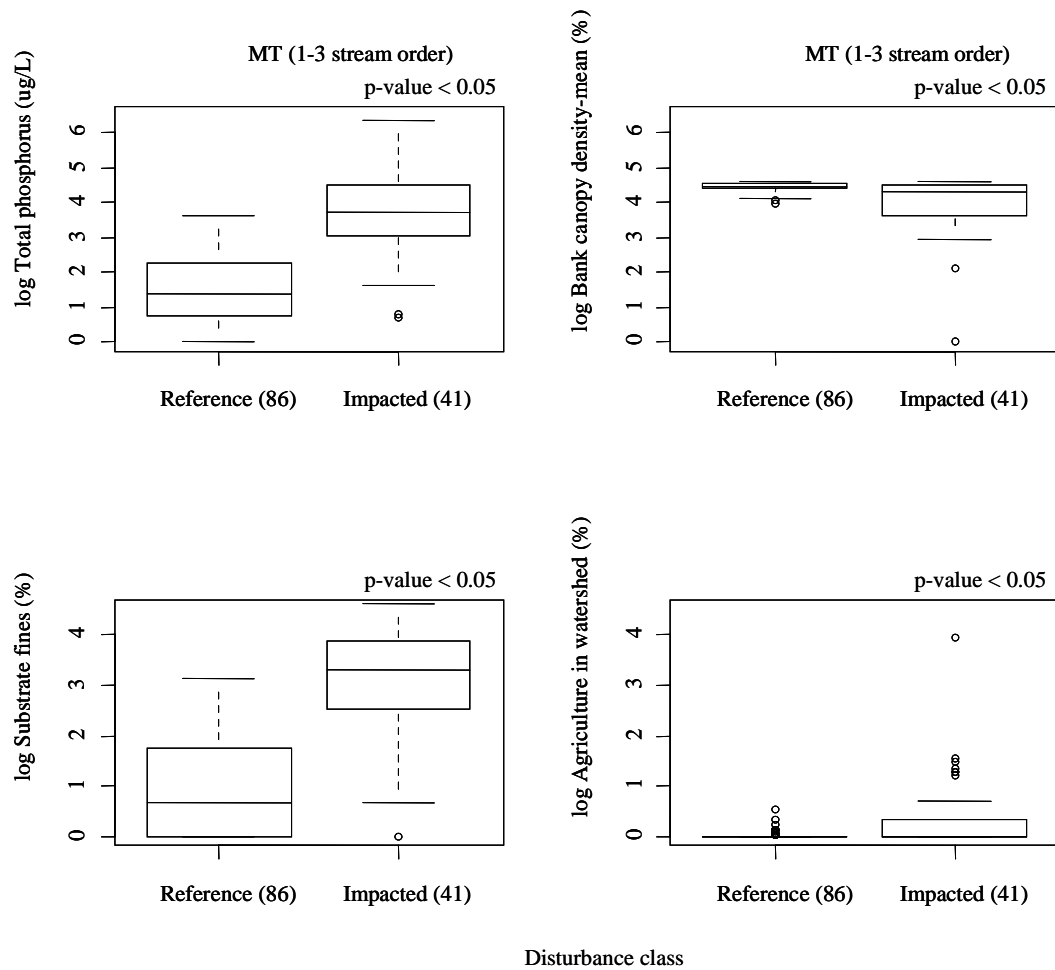


Figure 2.14. Boxplots for selected environmental variables from the conceptual model for reference and impacted sites (numbers in parentheses) in small (1<sup>st</sup>-3<sup>rd</sup> order) streams in the MT ecoregion.

## Discussion

Percentage live diatoms (34.5%) across Western US streams were relatively low compared to other studies which reported average values of 73% live diatoms from epilithic habitats (Peterson, 1987), 66.8% from estuarine sediments (Wilson and Holmes, 1981), and 63.4% from small mountain streams (Chapter 3). Potential explanations for this difference may include large heterogeneity among the analyzed streams (live diatom range 2.08-97.02%), time between sample collection and

processing (e. g., if not preserved well, cell contents start to decompose) or high abundance of other types of algae (e. g., both live diatoms and other algae contributed to the total number of counting units). There were pronounced distributional patterns of % live diatoms in Western US streams. They were decreasing from MT through XE to PL sites. In all three ecoregions, % live diatoms were also decreasing from small streams to larger ones. The observed relationships can be attributed to environmental features of each ecoregion and stream size within it. Distinctive differences in diatom assemblages between mountain and valley areas have been reported for Oregon (Whittier et al., 1988) and Mid-Atlantic Highlands (Pan et al., 2000). Despite the generally more stressful environmental conditions in the mountains (e. g., more disturbance by fast flow, less nutrients and light), % live diatoms there were higher than in more favorable environments in the plains (e. g., slow flow, more nutrients and light). However, sources of stress in the plains can originate from excess nutrients, high sedimentation and turbidity levels which decrease water clarity and light availability.

Seasonality may have affected both diatom species composition and % live diatoms. The mean Bray-Curtis similarity coefficient between two within-year site visits was 56.00% which is slightly lower than the recommended 60% for good similarity agreement (Kelly, 2001). However, its wide range (26.20-80.60%) reveals good similarity agreement for some sites, but not for others. Environmental characteristics of sites with low and high similarity coefficients do not provide a clue to potential causes for this variation. For example, the few sites with the highest

similarity (> 70%) had no agriculture in their watersheds while the site with the lowest similarity (26.17%) was sampled at 2 month interval. However, there were other sites which were sampled at such large intervals, but they did not have low similarity coefficients. Despite of the significant difference in % live diatoms between within-year site visits, there was a large amount of variability which may be due to a number of reasons including natural variability among samples collected from heterogeneous geographic regions (all 12 states). Assemblages may or may not change within a season, because additional factors (e. g., successional stage, resources, disturbance, and grazing) other than season alone may shape the periphyton community in a site (Peterson, 1987; Peterson et al., 1990; Stevenson et al., 1991). Generation times of benthic algae can be as short as two days which is much shorter than seasonal changes in their abiotic environment (McCormick and Stevenson, 1991). The rate of change in species composition can be faster or slower depending on light and nutrient availability (Stevenson et al., 1991), while grazing effects on algae may differ among years and seasons (Peterson et al., 2001).

Contrary to my expectations, % live diatoms tended to increase at late successional sites, but they were significantly higher only in the MT and small MT streams. The higher % live diatoms at late successional sites may be due to higher algal production and proliferation rates in these nutrient-rich and high-light habitats (Figs. 2.6, 2.8). Surprisingly, these habitats also had more substrate fines and higher turbidity, but smaller stream cross-section. These features are characteristic of the presence of logging activities which open up the canopy cover and increase nutrient,

turbidity and sediment levels (Naymik et al., 2005). The lower % live diatoms at early successional sites in the MT and small MT streams may have three potential causes: high grazing pressure, limited resources and high disturbance. The lack of significant difference in scrapers density between the two successional stages excludes the first possibility. Supportive evidence for the second and the third causes is provided by the fact that such low-nutrient, fast-water habitats are characterized with low immigration rates, high export and slow reproduction (Stevenson, 1996).

Contrary to my expectation, there were less grazers and % live diatoms at sites with piscivores. This finding is in agreement with Fretwell's prediction (1977, 1987) that in food webs with even number of trophic levels plants are grazer limited. As expected, % live diatoms and grazers density showed the same response to piscivore presence, they both decreased. There are two potential explanations for that: either % live diatoms limited grazers density and growth or vice versa (Feminella and Hawkins, 1995). The abundance of nutrients and light at these sites provides favorable conditions for algal growth and excludes the first possibility (Steinman et al., 1989; Stevenson et al., 1991). Therefore, grazers density may stimulate % live diatoms by increasing algal export and reducing competition pressure (Peterson, 1987; Lamberti et al., 1989), and by improving their resistance and resilience to disturbance (Mulholland et al., 1991; Steinman et al., 1991). However, the relationship between piscivores and grazers may not be that straightforward, because piscivores feed not only on invertivores, but on other fish (e. g., herbivores, other piscivores). The low herbivore richness in the dataset (maximum 1 individual/site) did not contribute much

to piscivores diet. Consumers at higher trophic levels are further away from the primary producers and may need more time to respond to changes in their prey (Carpenter et al., 1985). In addition, omnivory can be an important factor in freshwater ecosystems and omnivorous fish may not always be able to control their prey (Vadas, 1990).

My conceptual model about the factors controlling % live diatoms was partially supported by the data. The only two significant and negatively related variables with % live diatoms in Western US streams were stream cross-section and % agriculture in the watershed. Both of them point to an altitudinal gradient of increasing stream size and agriculture at low elevation. A similar “downstream” gradient in diatom distribution patterns was described from another large scale study in the US (Potapova and Charles, 2002). A decrease in % live diatoms with increasing nutrients, substrate fines and turbidity was observed at impacted sites in the PL (Fig. 2.11). I anticipated that only resources (light and nutrients) and grazers will stimulate % live diatoms. My results revealed that light, grazers and piscivore density consistently had positive effects on % live diatoms across spatial scales. Light increases algal biomass which provides food for grazers, while they clear space by increasing algal export (Lamberti et al., 1989; Steinman et al., 1989). However, nutrients, fine sediments, stream cross-section and agriculture in the watershed had variable effects (positive or negative) on % live diatoms depending on the spatial scale (ecoregion or stream size). The poor association between % live diatoms in Western US streams with biotic and abiotic factors from my conceptual model may partly be due to the large spatial

coverage of the dataset which resulted in a noisy relationship between % live diatoms and the selected variables. Landscape classifications usually do not explain large amounts of the variance in stream biota (Pan et al., 1999; Cao et al., 2007). Finer spatial scales (ecoregions and stream orders) helped increase the amount of explained variance in % live diatoms. Increasing % live diatoms in MT streams were related to decreasing stream cross-section (i. e., smaller streams) and increasing agriculture (i. e., more nutrients) and piscivore fish density (i. e., more scrapers which stimulate live diatoms). Percentage live diatoms in small MT streams were stimulated by substrate fines (i. e., source of nutrients) and piscivore fish density, but were suppressed by stream cross-section (i. e., high flow) and % agriculture (human disturbance). Percentage live diatoms in large MT streams were negatively affected only by substrate fines which may have decreased water transparency and light availability to photosynthesizing diatoms. Percentage live diatoms in xeric streams were negatively influenced by shade and substrate fines for the same reasons as in large MT streams. The highest amount of explained variation in % live diatoms was for small and large streams in the XE ecoregion, 21% and 55%, respectively. This may partly be due to the higher number of explanatory variables in these models or to better characterization of their relationship. Percentage live diatoms in small XE streams were negatively affected by total phosphorus, shading and substrate fines, but positively affected by stream cross-section. Percentage live diatoms in large XE streams were increasing with increasing concentrations of total phosphorus, stream cross-section, scrapers and piscivore density.

Percentage live diatoms were able to distinguish reference from impacted sites only in the MT and somewhat in the PL. The observed trend of increasing % live diatoms from reference to impacted sites in the MT and decreasing % live diatoms from reference to impacted sites in the PL may be due to two different sides of a unimodal relationship (Fig. 2.11). For instance, % live diatoms are stimulated to increase to a certain level after which they decrease as a result of some negative effect (an example of a subsidy-stress gradient, Odum et al., 1979). Support for this suggestion is provided by the mean total phosphorus concentrations at impacted sites in the MT (68.27  $\mu\text{g/L}$ ) and reference sites in the PL (62.19  $\mu\text{g/L}$ ). Reference sites in both ecoregions were forested and well-shaded compared to impacted sites which were nutrient-rich with more fine substrates and higher turbidity levels. While the latter conditions may promote % live diatoms in the Mountains, they may suppress them in the Plains. A similar difference between reference and impacted MT sites from the same dataset, but for vertebrate and invertebrate metrics, was found by Whittier et al. (2006). They also failed to find metrics which were able to distinguish reference from impacted sites in the PL. Whittier et al. (2006) concluded that reference sites in the PL are more disturbed than impacted sites in the MT, because of the pervasive human nature of land use. In addition, finding reference sites in the MT is much easier due to landscape limitations on human activity. Somewhat impacted sites in the PL may have been classified as reference to fulfill the target of 15% of all sites classified in this category (Whittier et al., 2006). In addition to the difficulty of finding reference sites in the PL, reference site designation in general is not unanimous, but



the project's probability-based sampling design provides objectivity about the selection criteria (Whittier et al., 2007).

### **Chapter 3: The importance of live diatoms in bioassessment: a regional study of Western US streams.**

#### **Introduction**

Accuracy and precision in diatom-based bioassessment are continually being refined to meet the increasing demands of water resource management. Recently, researchers have examined if sampling design (Potapova & Charles, 2005; Weihhoefer & Pan, 2007), lab analytical procedures (Alverson et al., 2003), and taxonomic consistency (Kelly, 2001; Prygiel et al., 2002) enhance the performance of diatoms in stream bioassessment. Unlike other bioindicators (macroinvertebrates and fish), conventional diatom bioassessment does not distinguish live (cells with protoplasts) from dead (empty cells) individuals. However, accounting for the amount of live diatoms in an assemblage may not differentiate reference from impacted conditions well (Chapter 2). Identification of live diatoms to fine taxonomic levels may further improve their signal about the environment. Such detailed information can not easily be collected for a large scale study like WEMAP which tries to find a balance between representative results and costs (Hughes and Peck, 2008), while the amount of effort for live diatom identification to species level is significant (Cox, 1998). Large scale studies such as WEMAP are good at detecting large-scale patterns and processes, but small scale studies may be better at uncovering local patterns. For example, diatom assemblages in streams are shaped by a variety of large scale (land use/cover in the watershed) and local factors (riparian condition) (Stevenson, 1997; Pan et al., 1999).

Therefore, a small study from a homogeneous area and species level live diatom identification will improve accuracy and precision of diatom-based stream bioassessment.

Our knowledge on the abundance of live and dead diatoms in aquatic systems is limited. A few studies were conducted from natural substrates, e. g. estuarine sediments (Wilson & Holmes, 1981; Oppenheim, 1987; Sawai, 2001; Hassan et al., 2008), plants (Thomas, 1979), stream rocks (Pryfogle, 1975; Pryfogle & Lowe, 1979; Owen et al., 1979), and artificial substrates (Patrick et al., 1954). However, they found contradictory information about the abundance of dead diatoms in an assemblage. Proportions of dead diatoms in depositional habitats were variable and high (Wilson & Holmes, 1981; Oppenheim, 1987). Sawai (2001) recognized two trends between live and dead diatoms in tidal wetlands: species numbers were higher for dead diatoms, and many dead diatom frustules/valves were transported widely. Patrick et al. (1954) concluded that dead diatoms were rare on artificial substrates in streams, while others found that more than 50% of the diatoms in a sample were dead (Pryfogle & Lowe, 1979; Owen et al., 1979; Wilson & Holmes, 1981).

It is unclear whether differentiating live and dead diatoms in an assemblage would contribute significantly to the accuracy and precision of diatom-based bioassessment. The potential errors in diatom-based bioassessment when analyzing the entire diatom community have been discussed by a number of scientists (Cox, 1998; Round, 1998), but without the support of empirical data. Several researchers have suggested that assemblages generated by the traditional method may better integrate

spatial and temporal variability of the sampled stream and its watershed, because the accumulation of live and dead diatoms over space and time may provide broad environmental information (Cox, 1998; Stevenson & Pan, 1999). However, Pryfogle & Lowe (1979) concluded that the traditional counting technique might mask the specific response of the resident taxa to the local environmental conditions and introduce more noise to the data by its inability to distinguish dead diatoms. The quantitative interpretation of the information embedded in dead diatoms in an assemblage is still unclear. It has been discussed that the contamination with dead diatoms from adjacent habitats might be a significant source of error in estimations of diatom communities and their signals about the local environment (Round, 1998). The advantages of live diatoms for bioassessment have been emphasized by Cox (1998), but without evidence for their data quality. She argued that live diatom identification will speed up and ease the process of bioassessment, decision making and management. However, there are no comprehensive floras based on live diatom cell morphology (except Cox, 1996) which hinders their identification and use in bioassessment. Therefore, a study that would compare the response of the whole diatom assemblage and its live component to the measured environmental variables would benefit diatom-based stream bioassessment.

In this study I compare the diatom assemblages generated by counting the traditional acid cleaned samples, which do not allow separation of live from dead diatoms, to the live diatoms from the undigested samples. Specifically, I want to quantify the proportion of live individuals in an assemblage and assess if there is a

correspondence in terms of species abundance with the diatoms of the whole assemblage. I anticipate that there will be mostly live diatoms on the sampled rock surfaces, and that their species composition will be very similar to the one of the whole assemblage. Next, I want to examine if the assemblages from the two counts differ in their responses to the measured environmental variables. Finally, I would like to see which environmental variables may account for the differences, if there are any.

## **Material and Methods**

### *Study Area and Site Selection*

Stream sites were located in the Northern Oregon Coast which is part of the Coast Range ecoregion (Level III ecoregions; Omernik, 1987). It is underlain by marine tuffaceous sandstones and shales with basaltic volcanic rocks (Jackson & Kimerling, 1993). The Northern Oregon Coast is characterized by high precipitation and maritime climate with cool, dry summers, and warm, wet winters (Naiman & Bilby, 1998). The Coastal mountains receive most of their precipitation, which averages between 203-254 cm a year, between October and March or April (Jackson & Kimerling, 1993). Peak stream flows occur in late fall and winter. It is a forested region dominated by western hemlock (*Tsuga heterophylla* (Raf.) Sarg.), Sitka spruce (*Picea sitchensis* (Bong.) Carr.), Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) and red alder (*Alnus rubra* Bong.) (Clarke et al., 1991). Land use in the area is mainly ungrazed forest and woodland (Jackson & Kimerling, 1993). Streams of the Coast Range ecoregion are naturally cool, clear, typically shaded and oligotrophic in nutrient

status. The sampling sites were selected based on geology, land cover and land ownership, and stream order (1<sup>st</sup> and 2<sup>nd</sup> order streams). From the initial pool of prospective sites, only watersheds dominated by volcanic bedrock lithology and well-forested were retained. All sites were on public lands and allowed access for sampling.

#### *Watershed characterization*

Watershed boundaries above each sampling reach were delineated with the Spatial Analyst extension of ArcInfo using 10-m digital elevation models from USGS (CLAMS, 1996). Bedrock lithology categories within each watershed were quantified from the geological map of Oregon (USGS 1:500,000, [www.gis.oregon.gov](http://www.gis.oregon.gov); Walker & Macleod, 1991). Geologic types were grouped in two major classes (volcanic and sedimentary). Vegetation within each watershed was quantified from the Gradient Nearest Neighbor Vegetation Class developed by CLAMS (1996). Vegetation cover within the watersheds was grouped in three major categories: broadleaf, conifer and mixed forest (broadleaf and conifer). Ownership data were also obtained from CLAMS (1996). ArcGIS (version 9, ESRI Inc. 2006) was used for all spatial analysis and calculations.

#### *Field Sampling*

Diatom samples were collected from 25 stream sites in the Coast Range of Oregon in September 2006 (Fig. 3.1). The sampling procedures followed a modified US EPA's Environmental Monitoring and Assessment Program (EMAP) protocol for wadeable rivers and streams (Lazorchak et al., 1998). The stream reach, 100 m in length, was divided into 9 equally-spaced transects. Along each transect, one diatom

sample was collected at  $\frac{1}{4}$  (left),  $\frac{1}{2}$  (center), or  $\frac{3}{4}$  (right) of the distance from the stream bank after a random start. One rock from each of the nine sampling transects was removed from the stream, an 8-cm<sup>2</sup> surface area was scraped with a toothbrush and washed into a collection bottle with stream water. The scrapings from the 9 rocks were combined into one composite sample which was frozen until the time of processing.

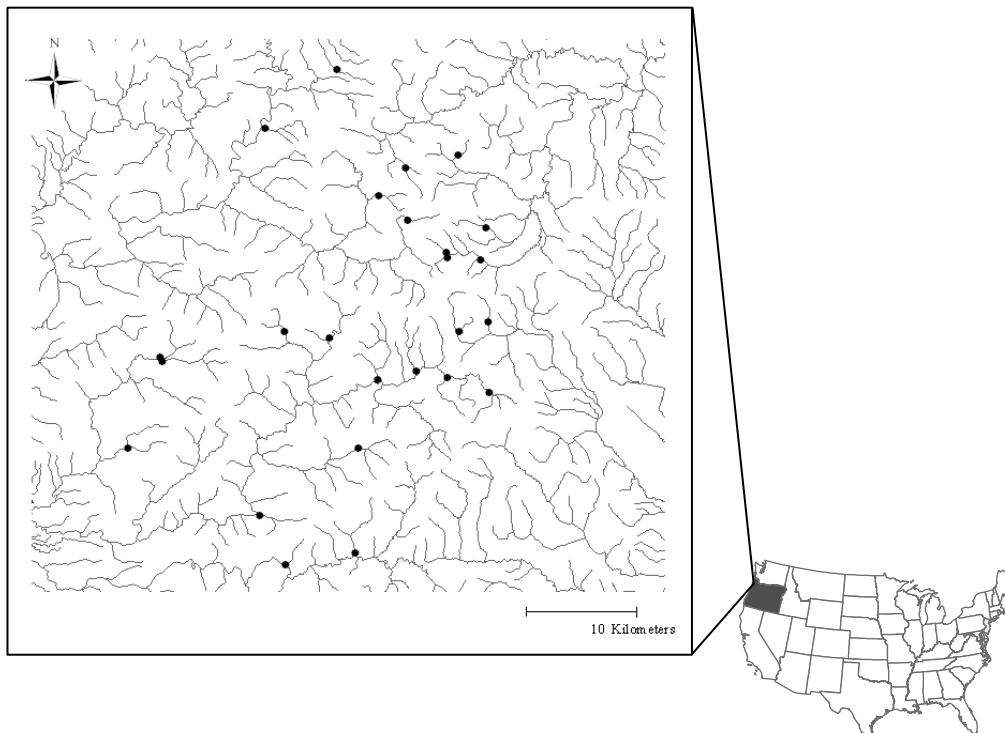


Figure 3.1. Map and location of the sampling sites in the Northern Oregon Coast Range, OR, USA.

The water quality variables (e. g. pH, temperature, salinity, dissolved oxygen, and specific conductivity) were measured in the middle of the stream reach with YSI 556 MPS. Unfiltered water samples were taken from the middle of the stream for total

Kjeldahl nitrogen and total phosphorus analysis. The physical habitat conditions were characterized at 5 points along each transect (two banks and 25 %, 50%, 75% wetted width). At each point, stream depth, substrate size and type were recorded. Substrate size and type for the reach were expressed as percent coverage following Kaufmann et al. (1999). Stream width, canopy cover and riparian vegetation along each transect were measured also. The stream depth and width were represented by their mean values for the reach. The canopy cover was measured with a spherical densiometer in the middle of the stream facing upstream, downstream, the left and the right banks. The riparian vegetation, estimated as vegetation density and type for three layers (canopy, understory and ground cover), was evaluated for each stream bank stretching 5 m upstream and 5 m downstream from the transect, and 10 m inland from the bank, and was expressed as percent coverage (Kaufmann et al., 1999).

#### *Lab Analysis*

Water nutrients (total Kjeldahl nitrogen and total phosphorus) were analyzed according to EPA methods 351.2 and 365.4 (US EPA, 1983). Each composite algal sample was split in two. The first split was used for chlorophyll *a* analysis following standard methods (Clesceri et al., 1998). Chlorophyll *a* concentrations in 90% aqueous acetone before and after acidification with 0.1N HCl were determined on Beckman DU-640 spectrophotometer. The second sample split was preserved with 40% formalin and used for composition analysis. Ten ml of this split were digested with HNO<sub>3</sub> using a MARS™ 5 microwave for conventional diatom analysis (Charles et al., 2002). After repeated rinsing with distilled water, clean frustules were dried on a glass



coverslip, and mounted on a glass slide using Z-RAX<sup>®</sup> mounting medium. A minimum of 600 diatom valves were identified from this slide, and because no distinction between live and dead diatoms could be made it was termed LD count (Live + Dead). The remaining volume of the sample split was used for the making of a wet mount. It was prepared in the following way, a few drops of the undigested sample were placed on a glass coverslip which was air dried, inverted and mounted in a drop of water on a glass slide. The coverslip was sealed with clear nail polish to prevent evaporation and allow the use of oil immersion. The wet mounts were used for the identification and enumeration of live and dead diatoms. The live diatoms were defined as the ones with visible cell contents while the dead diatoms were the ones with empty frustules. To be consistent with the LD count, three hundred live diatom cells (L count) were identified and counted. During the analysis of live diatoms, dead diatoms (D count) were also identified and enumerated. The L and D counts from the wet mount were used to calculate the percentages of live and dead diatoms. All diatoms were identified to the lowest possible taxonomic level (mainly species) at 1000x magnification, using Leica DM LB2 light microscope with differential interference contrast (DIC). The diatom taxonomy followed predominantly Krammer & Lange-Bertalot (1986, 1988, 1991a, b) and Patrick & Reimer (1966, 1975).

#### *Data Analysis*

Environmental variables which did not approximate normal distributions were log<sub>10</sub> or square root transformed. Diatom counts were represented as relative abundances. To compare the three assemblages (LD, L and D) I looked at: species

richness, relative abundance of dominant species, similarity between samples and site ordinations. Species richness between the samples and the relative abundances of the dominant species between the three different assemblages (LD, L and D) were evaluated with paired t-tests due to their dependent nature. The similarity among the three assemblages for each stream site was assessed using Bray-Curtis similarity coefficient, after square-root transformation of the data to down weight the effect of dominant species. Bray-Curtis similarity values range between 0 (complete difference) and 100 (complete similarity). This coefficient takes into account both species presence and abundance, and is commonly used in analysis of ecological communities (Clarke, 1993). To evaluate whether the similarity between LD and L counts is related to the species richness or percent live diatoms, a simple linear regression was performed. The inter-site similarities were used in non-metric multidimensional scaling ordination (NMDS) to project their relationships into a low-dimensional space and to best preserve the ranked distances between them. NMDS does not require any assumptions about the species distribution, and allows for user-specified distance measure. To assess if the inter-site relationships defined by their similarity coefficients were well projected into the NMDS plot, the stress value was calculated. It shows how good the correspondence between the calculated (plotted) and the actual distances (from the similarity matrix) between the sites is, where a lower value indicates a better ordination. NMDS was run separately for LD, L and D assemblages. The resulting three NMDS plots were then compared using procrustes analysis with permutation test (Gower, 1971; Peres-Neto & Jackson, 2001). Procrustes analysis

rotates two ordinations to maximize the similarities between them by minimizing the sum of their squared distances ( $m^2$  statistic, Gower, 1971). Higher values of the  $m^2$  statistic indicate better correspondence between two ordinations. Unlike Mantel test, procrustes analysis allows for a site-by-site comparison by returning pairwise residuals (Peres-Neto & Jackson, 2001). The significance of overall similarity between any two NMDSs was assessed using protest, a permutation test (Jackson, 1995). It evaluates whether the degree of resemblance between two ordinations is greater than random. To explore the relationship of each assemblage with its environment, a linear fitting function was then used. This function finds vector averages of the environmental variables, and fits them in the ordination space defined by the species data. The significance of each vector was assessed with a goodness-of-fit statistic (squared correlation coefficient,  $r^2$ ) using 1000 permutations. To further examine which environmental variables might be responsible for the differences in the comparisons between the three ordinations (LD, L and D), I calculated the Pearson correlation coefficients between the residuals from the procrustes analysis and the environmental variables (Peres-Neto & Jackson, 2001). All data analyses were performed using the ‘vegan’ (Oksanen et al., 2008) and ‘MASS’ (Venables & Ripley, 2002) packages in R (R Development Core Team, 2008).

## **Results**

The sampled stream reaches were generally small and well-shaded (Table 3.1). Most streams were narrower than 7.0 m, but their stream width ranged from 1.9 to

10.1 m. The streams were shallow, with a mean depth of 4.0 to 14.8 cm. Mid-channel canopy cover ranged between 50-100% with an average of 90%. The water temperature ranged from 9.5 to 13.0°C, and pH was slightly acidic to neutral (range 5.5 to 7.5). All streams had very low nutrient levels. The mean value of chlorophyll *a* was 0.025 mg/m<sup>2</sup>. Watershed area ranged between 2.4 and 23.7 km<sup>2</sup>.

Table 3.1. Minimum, mean and maximum values for some environmental variables for Oregon Coast Range streams (n=25).

	Min	Mean	Max
<b>Algal biomass</b>			
Chlorophyll <i>a</i> (mg/m <sup>2</sup> )	0.01	0.03	0.06
<b>Water quality</b>			
pH	5.5	6.9	7.5
Temperature (C°)	9.5	11.2	13.1
Specific conductivity (µS/cm)	41.0	68.2	203.0
Total phosphorus (mg/L)	0.01	0.02	0.05
Total Kjeldahl nitrogen (mg/L)	0.01	0.03	0.07
<b>Physical habitat</b>			
Watershed area (km <sup>2</sup> )	2.4	9.8	23.7
Mean depth (cm)	4.1	8.7	14.8
Mean wetted width (m)	1.9	4.6	10.2
Canopy cover (%)	50.0	89.5	99.7
Riparian deciduous canopy (%)	0.0	70.4	100.0
Riffles (%)	33.3	70.7	88.9
Smooth bedrock (%)	0.0	6.4	31.1
Fine gravel (%)	0.0	9.2	28.9
Sand (%)	0.0	3.2	13.6
Fines (Silt/Clay/Muck) (%)	0.0	3.5	37.8

A total of 135 species were recorded in the LD counts, with a mean species richness of 26 (Table 3.2). Fewer taxa (90) were identified in the L counts, with mean species richness of 19. The D counts were slightly more diverse than the L ones with a total of 96 species, and a mean species richness of 20. The species richness of LD was significantly different from the other two assemblages (L and D, Fig. 3.2A-B). There

were 47, 7 and 12 species unique to the LD, L and D counts, respectively, occurring with less than 1% mean relative abundance. Most species encountered were periphytic in their habitat preferences. There were only a few euplanktonic species occurring with relative abundances below 1%. The paired t-tests revealed some interesting patterns in the differences in relative abundances of the dominant species between the LD, L and D assemblages. Some species had significantly different relative abundances between LD and L, where they were either over- or underestimated (Fig. 3.3). A slightly smaller number of dominant species were found to differ significantly in their relative abundances between D and L (Fig. 3.4).

Table 3.2. Diatom community characteristics with mean (minimum and maximum) values.

	<b>LD</b>	<b>L</b>	<b>D</b>
Total number of species	135	90	96
Sample richness	26 (13-50)	19 (10-42)	20 (11-40)
Dominant species (% relative abundance)			
<i>Achnantheidium minutissimum</i> (Kützing)			
Czarnecki	23 (1.5-85)	28 (3.7-89)	25.1 (2.3-76.6)
<i>Rhoicosphenia abbreviata</i> (Agardh)			
Lange-Bertalot	16.3 (0-57.2)	12.1 (0-56)	8.7 (0-27.2)
<i>Achnantheidium deflexum</i> (Reimer)			
Kingston	9 (0.3-43)	11.6 (0.5-50)	9.4 (0.4-43.3)
<i>Planothidium lanceolatum</i> (Brébisson)			
Lange-Bertalot	8.2 (0-26)	5.6 (0-25)	9.3 (0-29.5)
<i>Cocconeis placentula</i> Ehrenberg	8 (0.5-40)	4.6 (0-23)	8.7 (0.3-40)
<i>Gomphonema pumilum</i> (Grunow)			
Reichardt et Lange-Bertalot	7.8 (0.8-40.2)	7.1 (0.3-50.5)	8.5 (1.6-37.2)
<i>Nitzschia inconspicua</i> Grunow	5.2 (0-32)	7 (0-50)	7.1 (0-43)
Mean % dominance by 1 taxon	39.8	43.7	36.5

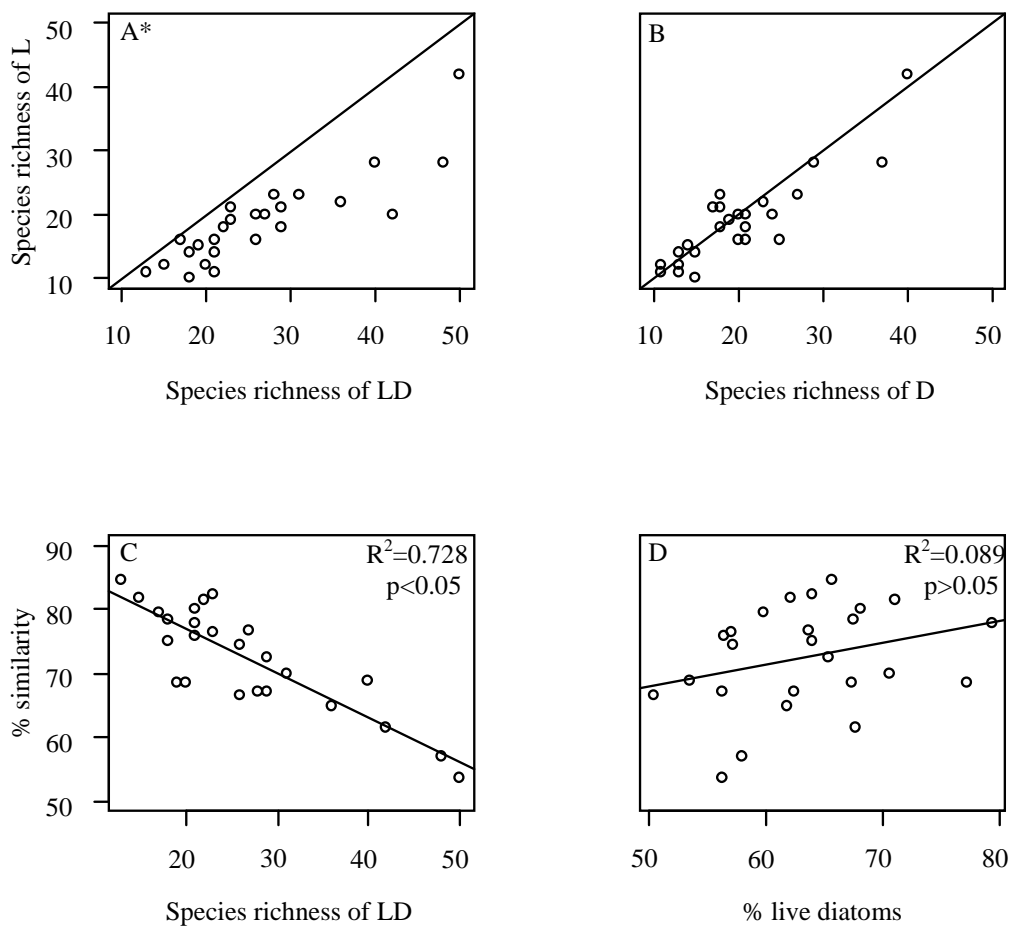


Figure 3.2. Scatterplots for different diatom community characteristics: (A) species richness of LD and L, (B) species richness of D and L, (C) species richness of LD and percent similarity between LD and L, (D) percent live diatoms and percent similarity between LD and L. The lines in A and B are 1:1 ratio. \* significance at  $p<0.05$ .

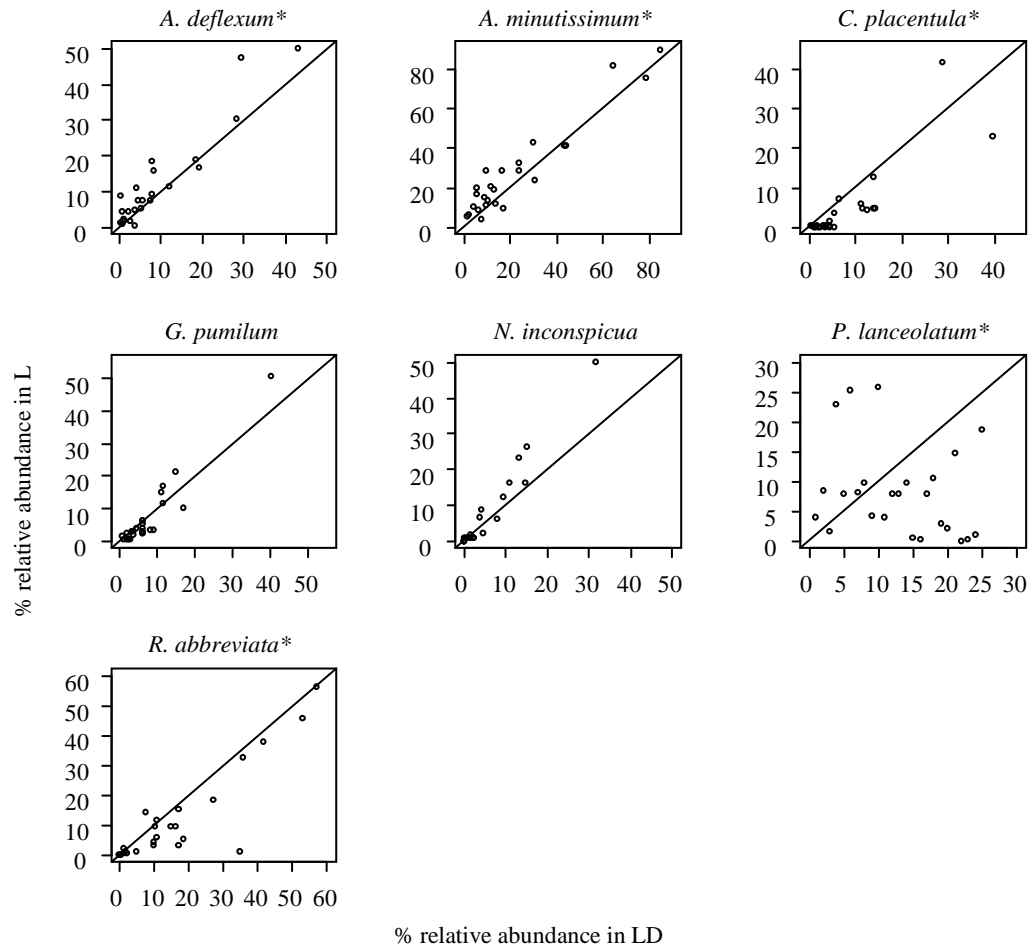


Figure 3.3. Scatterplots for the percent relative abundances of the dominant diatom species in LD and L assemblages with a 1:1 ratio line. Refer to Table 2 for complete species names. \* significance at  $p < 0.05$ .

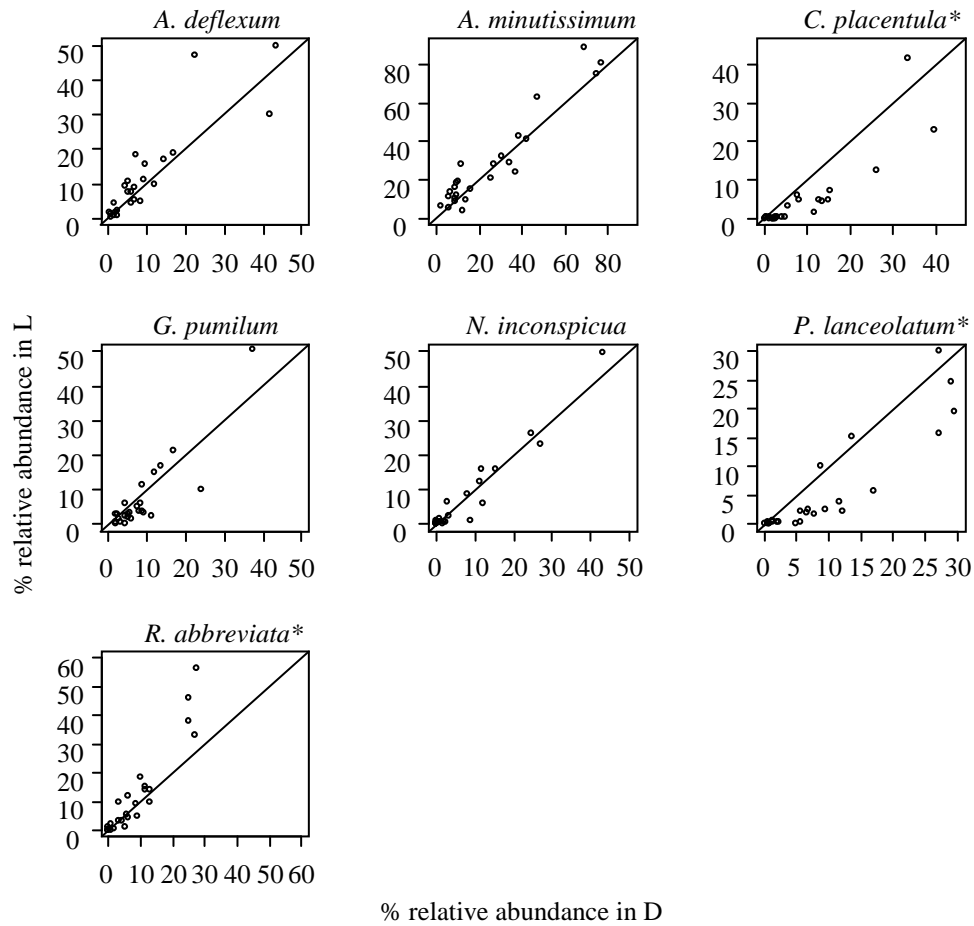


Figure 3.4. Scatterplots for the percent relative abundances of the dominant diatom species in D and L assemblages with a 1:1 ratio line. Refer to Table 2 for complete species names. \* significance at  $p < 0.05$ .

The mean percentage of live diatoms in a sample was 63.4% (range 50.5-79.4). The assemblages were overall similar. Their mean similarity values were 72.4%, 73.4% and 75.4% for LD and L, L and D, and LD and D, respectively. The percentage of live diatoms did not exhibit a statistically significant relationship with the similarity among samples (Fig. 3.2D). The sites with the lowest similarity had the highest species richness in LD assemblages, while the sites with the highest similarity coefficients had some of the lowest species richness (Fig. 3.2C). The low similarity



values were primarily due to many rare species (species with low abundances), while the high similarity coefficients were due to many common species. The two-dimensional NMDS plots for the LD, L and D assemblages had stress values of 16.0, 16.7 and 17.0, respectively. The site with the highest similarity coefficient was dominated by *Achnanthidium minutissimum*, which reached its highest relative abundance here (85%, Fig. 3.5A). Only a few of the dominant species showed pronounced distributional patterns in their relative abundances among the sites, e. g. *A. minutissimum* had a similar pattern of increasing relative abundance along the 1<sup>st</sup> NMDS axis for both LD and L ordinations (Fig. 3.5).

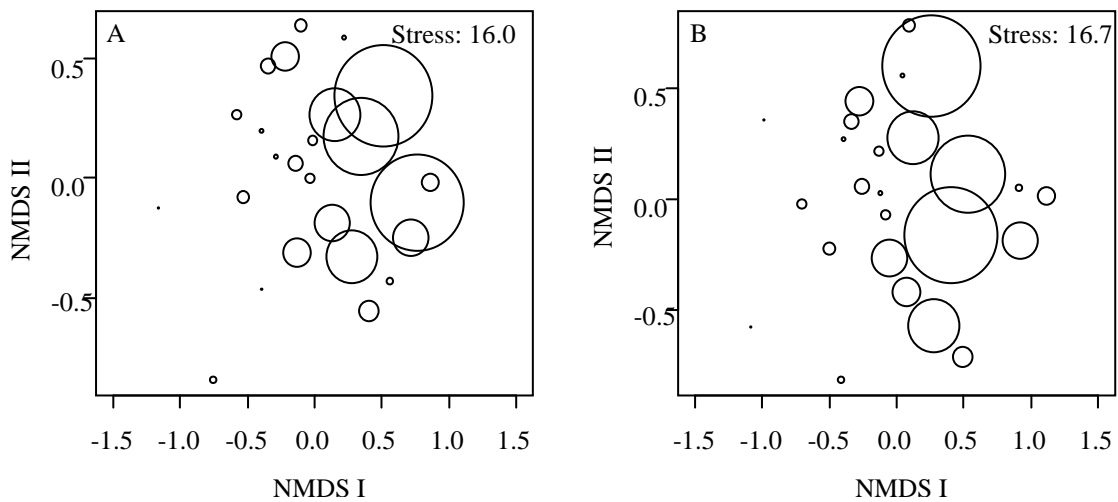


Figure 3.5. Bubble plot for the distribution of relative abundance of *Achnanthidium minutissimum* among the sites for the LD (A) and L (B) counts. Bigger bubble size corresponds to higher relative abundance.

The comparisons between the site ordinations were summarized in plots of the residuals for the procrustes rotations. The plot for the LD and L ordinations is presented on Fig. 3.6. Deviation in the species composition between LD and L count

for a site is indicated by an arrow connecting the two counts. The longer the arrow, the greater the difference between the two counts. Some sites had larger differences in their species composition than others. The configuration for the procrustes rotation of the NMDS plots for LD and L counts, as well as the one for LD and D counts, matched well ( $m^2 = 0.85$ ,  $p < 0.001$ , 1000 permutations). The configuration for the L and D ordinations was even better ( $m^2 = 0.95$ ,  $p < 0.001$ , 1000 permutations).

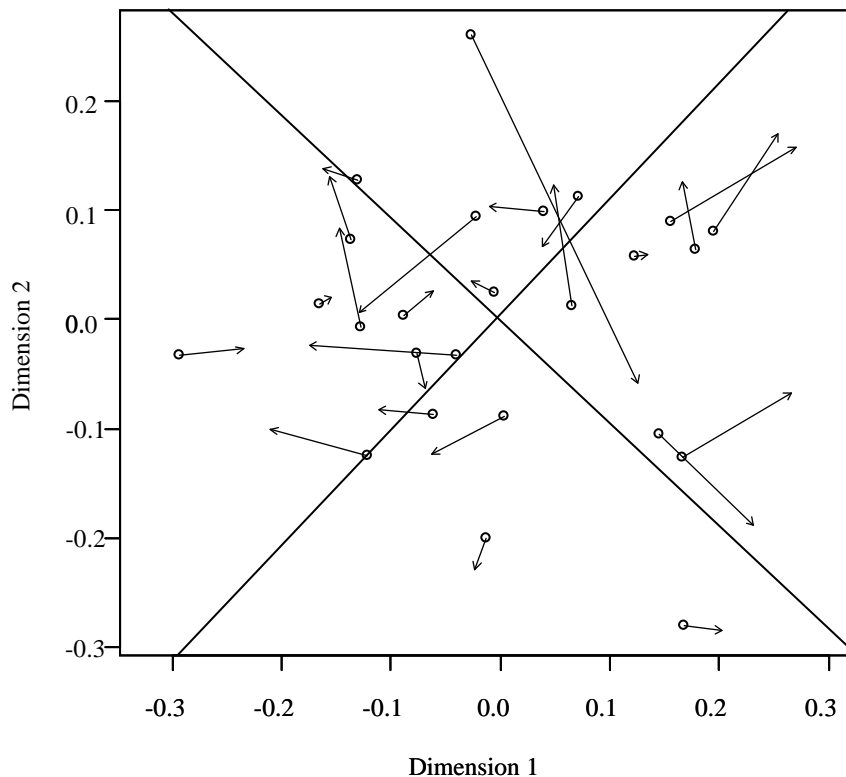


Figure 3.6. Plot of the residuals from procrustes analysis for the comparison between the ordinations of LD and L. Each arrow connects the LD and the L count from one site. The arrows point in the direction of the L ordination. The arrow length corresponds to the difference in site location between the two ordinations.

Seven environmental variables co-varied significantly with the species data ( $p \leq 0.05$ , Table 3.3). These variables defined a gradient of primary production and

physical habitat conditions (e.g., channel dimensions, substrate types, and canopy cover).

Table 3.3. Squared correlation coefficient ( $r^2$ ) and its significance between the selected environmental variables and the three assemblages (LD, L and D).

	<b>LD</b>		<b>L</b>		<b>D</b>	
	$r^2$	p-values	$r^2$	p-values	$r^2$	p-values
<b>Algal biomass</b>						
Chlorophyll <i>a</i> (mg/m <sup>2</sup> )	0.496	<0.001***	0.367	0.008**	0.394	0.003**
<b>Water quality</b>						
Specific conductivity (µS/cm)^	0.145	0.164	0.109	0.281	0.157	0.152
Total phosphorus (mg/L)	0.034	0.682	0.120	0.228	0.098	0.310
Total Kjeldahl nitrogen (mg/L)	0.095	0.343	0.116	0.266	0.083	0.378
<b>Physical habitat</b>						
Canopy cover (%)	0.352	0.009**	0.404	0.006**	0.342	0.016*
Mean velocity (m/s)	0.135	0.188	0.087	0.366	0.244	0.051.
Mean wetted width (m)^	0.491	0.001***	0.491	0.001***	0.375	0.003**
Riffles (%)^	0.091	0.364	0.094	0.340	0.174	0.109
Smooth bedrock (%)^	0.226	0.061.	0.255	0.038*	0.127	0.214
Rough bedrock (%)	0.010	0.892	0.021	0.792	0.009	0.904
Large boulders (%)	0.087	0.379	0.170	0.120	0.109	0.262
Fine gravel (%)^	0.027	0.712	0.009	0.901	0.077	0.395
Sand (%)	0.305	0.018*	0.182	0.112	0.100	0.328
Fines (Silt/Clay/Muck) (%)^	0.478	0.001***	0.426	0.002**	0.495	<0.001***
Riparian deciduous canopy (%)	0.340	0.015*	0.114	0.278	0.203	0.072.

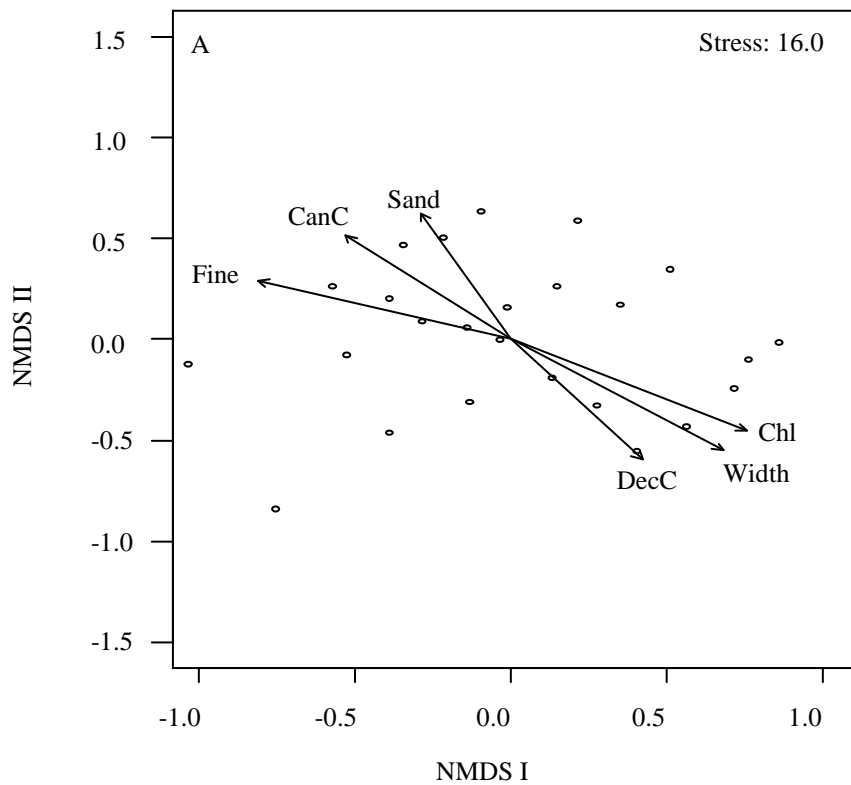
p-value codes: 0 '\*\*\*\*' 0.001 '\*\*\*' 0.01 '\*\*' 0.05 '\*' 0.1 '.'

p-values based on 1000 permutations

^ transformed variables

There was a change from bigger streams with deciduous riparian vegetation and higher primary production to smaller less productive shaded streams with finer substrates (Fig. 3.7). Four variables (chlorophyll *a*, canopy cover, stream width and fine sediments) were significant for all three assemblages. The other three variables (smooth bedrock, percent sand and riparian deciduous canopy) had varying influence

on one or more of the assemblages. The Pearson correlation coefficients between the residuals from the procrustes analysis and the environmental variables showed which variables were responsible for the differences in the comparisons of the three ordinations (LD, L and D). The only significant correlation was between chlorophyll *a* and the residuals from the procrustes rotation of LD and L.



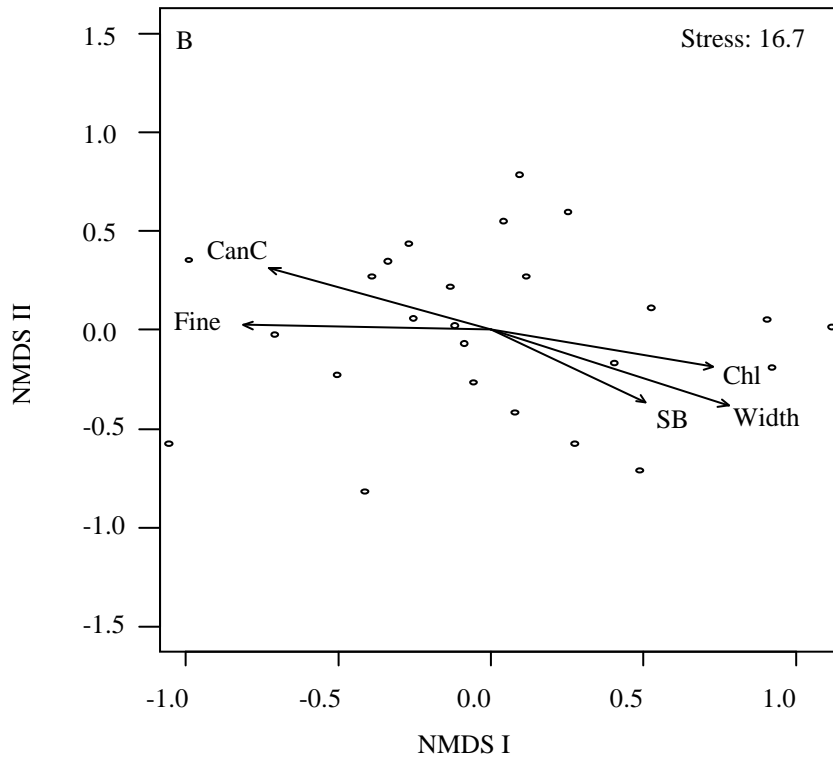


Figure 3.7. Non-metric multidimensional scaling plots for the LD (A) and L (B) assemblages, and significant environmental variables ( $p < 0.05$ ) from the linear fitting function. Abbreviations: CanC – canopy cover, Chl – chlorophyll *a*, DecC – percent riparian deciduous canopy, Fines – percent fine substrates (silt/clay/muck), Sand – percent sandy substrates, SB – percent substrates as smooth bedrock, Width – mean wetted width.

## Discussion

The traditional acid-cleaned diatom assemblage and the unprocessed one were overall similar. Pair-wise comparisons showed that the species composition generated by the two methods was overall similar (mean Bray-Curtis similarity = 72.4%). NMDS plots suggested that the overall relationships among stream sites based on their ranked Bray-Curtis similarity coefficients were similar. A more quantitative evaluation of the match between the two ordinations was achieved with the procrustes rotation which showed that there was a good agreement between the two ordination

configurations ( $m^2 = 0.85$ ,  $p < 0.001$ , 1000 permutations). In addition, both assemblages responded to a similar set of environmental conditions. In this sense, these results are consistent with several other studies that found similar species composition between the two assemblages in creeks, tidal wetlands and estuaries (Pryfogle, 1975; Sawai, 2001; Hassan et al., 2008). Hassan et al. (2008) also concluded that the living diatom assemblage at the estuarine sediment surface was very similar to the subsurface one, and that their ecological preferences could be used to infer past changes in their environment. This study reinforces the statement that the similarity between the two assemblages (LD and L) and the similarity in their relationships with the local environment justify their individual use for bioassessment purposes, because they will both provide similar results. We do not necessarily need to simultaneously identify live diatoms and acid-cleaned ones, but we can use each one of them independently.

Diatom analysis using the traditional counting method provides assemblage data with high taxonomic resolution and taxa richness. Both of them are of key importance in bioassessment. For instance, the multimetric (MMI) and multivariate (e.g., RIVPACS) indices commonly used in stream bioassessment, are based on species composition (Chessman et al., 1999; Hawkins, 2006; Cao et al., 2007). The finer taxonomic resolution facilitates bioassessment where species level identification improves the accuracy of prediction models because species have narrower environmental requirements than genera (Chessman et al., 1999; Hawkins, 2006). This study shows that the LD had 33% higher total species richness and 27% higher

average species richness than those generated in L (Table 3.2). Similar results have been reported by Wilson & Holmes (1981) who counted 300 frustules of live and dead diatoms to approximate the diversity of a cleaned assemblage and to calculate live/dead ratios. They found that measurements of community richness and diversity increased when dead diatoms were included in their calculation. However, it is difficult to attribute the difference in species richness between the two assemblages to inability to identify properly all live diatom taxa or to recognize if a large proportion of rare taxa in a benthic diatom assemblage were dead. It is a common practice in diatom analysis to identify and enumerate species from the LD count and then estimate total live diatom density by counting but not identifying all live diatoms from the wet mount or from a Palmer-Maloney counting cell (Charles et al., 2002). To estimate the density of each diatom taxon, its relative abundance from the LD count is multiplied by the total number of live diatoms from the wet mount or from the Palmer-Maloney counting cell. Potapova & Charles (2005) suggested that this extrapolation might overestimate live diatom species richness. This is in agreement with my results that L counts overly underestimated species richness (Fig. 3.2).

Technical difficulties may hinder the utilization of live diatoms in benthic bioassessment, despite of the potential valuable information embedded in them. They are more difficult to identify because the protoplast obscures the cell wall morphology. Analysts still need to examine cleaned slides using the traditional method to aid identification of live diatoms. In addition, the lack of floras or keys for their identification requires every analyst to make their own and confirm questionable live

taxa with cleaned material (Cox, 1998). As a result of their ambiguous identification, analysts might unintentionally lump similar species together and reduce taxa richness which is crucial for many bioassessment methods (e. g. many live *Navicula*'s are difficult to recognize, Thomas, 1979). Distinguishing live diatoms might be important in ecological studies where cell density, biovolume and growth rate need to be estimated (e. g. experimental studies, Barnese & Lowe, 1992; Peterson, 1996) or to encourage non-specialists to recognize dominant or ecologically informative genera (Cox, 1998).

The assumption of proportional abundance for live and dead diatoms in a benthic diatom assemblage may not be valid for all species (Fig. 3.3-3.4). The discrepancy will affect bioassessment methods which heavily rely on good estimation of species abundance, e. g. weighted averaging, when the living component of the assemblages better reflects environmental conditions such as nutrients. The high numbers of dead diatoms in the samples contributed to the differences in the species composition between LD and L. Owen et al. (1979) reported that only a few of their samples from natural and artificial substrates contained less than 10% empty frustules. In my samples, the lowest abundance of dead diatoms was 20.6%. Such high numbers of dead diatoms were reported in other studies which found that half of the diatoms collected in the spring and summer were dead while for the rest of the year these same numbers were somewhat lower (Pryfogle & Lowe, 1979). Owen et al. (1979) observed high numbers of dead diatoms on the glass slides regardless of the period of their immersion. Only 12.5% of the dead species in this study had relative abundance above



1%. Owen et al. (1979) noticed that less abundant diatoms were often represented by empty frustules which have been observed for months after the species abundance have declined. The inclusion of dead diatoms might introduce some error to the data and its subsequent analysis by incorporating non-resident species. My research revealed that the sampled erosional habitats, e. g. riffles with epilithic communities, accumulated high numbers of dead diatoms, but they were mainly resident to it.

## **Summary, Implications and Recommendations for Water Quality Assessment**

Diatom-based stream bioassessment is constantly being improved to meet the increasing demands of water quality management. This dissertation research examined whether the number of rare taxa and percentage (%) live diatoms can be used as indicators of human disturbance in streams/rivers. My findings reveal that rural taxa richness can be used to distinguish reference from impacted sites, but only in the Mountains ecoregion (t-test,  $p < 0.05$ ) (Chapter 1). However, they do not distinguish sites so well. There was an overlapping variation in the numbers of rural taxa at reference and impacted sites even in the Mountains ecoregion. Therefore, rarity definition is important for bioassessment, because other more appealing metrics based on satellite and urban taxa which may be considered good indicators of pollution, do not work well. If we are interested in specific rare taxa, the best approach to find them is to sample similar environments and look for them there (Finlay et al., 2002). Some previously considered endemic species from Lake Baikal were also found in the neighboring regions (Genkal and Bondarenko, 2006). If we are interested in using rare taxa as a bioassessment metric, we may have to modify the project design. Rare species detection may be improved by sampling larger areas (Cao et al., 1997, 2002), specific species rich habitats (Soininen and Eloranta, 2004), or conducting long diatom counts (Patrick et al., 1954). Sampling all available habitats in large heterogeneous stream reaches may be necessary to characterize its diverse diatom assemblages (Weilhoefer and Pan, 2007). This may require modification of the current equidistant transects location which may miss habitats. Fixed-count methods may work well to

capture macroinvertebrates species richness (Barbour and Gerritsen, 1996), but it may not capture the diatom assemblage well (Pappas and Stoermer, 1996). In unpolluted streams, 8000 diatom specimens may need to be counted (Patrick et al., 1954), although the true species richness in a site may never be estimated (Mao and Colwell, 2005). Also, stratified counting can be employed to better estimate the diatom community (Stevenson and Pan, 1999). For instance, the counting of dominant taxa can stop when a predefined number of cells is reached while the counting of rare taxa can continue until a prespecified number of individuals is encountered. The presence of rare taxa may be misleading if no distinction is made between living and dead diatom cells, because this may result in overestimated species richness (Pryfogle & Lowe, 1979; Wilson and Holmes, 1981). Without such distinction it will be impossible to tell whether a taxon was growing in a site at the time of sampling or whether it was washed in as a dead individual (Round, 1998). More than 50% of the diatoms in a sample may be dead, especially in depositional habitats (Wilson & Holmes, 1981; Oppenheim, 1987).

Percentage live diatoms was able to distinguish reference from impacted sites only in the Mountains ecoregion (t-test,  $p=0.02$ ) and somewhat in the Plains (t-test,  $p=0.05$ ) (Chapter 2). However, % live increased with disturbance in the Mountains and decreased in the Plains. Therefore, caution should be exercised to separately analyze samples from the two regions. Percentage live diatoms would be a time and cost effective bioassessment tool. They require only recognition and enumeration of live (cells with protoplast) and dead (empty) diatom cells, and no sample processing or

species level identification. However, the significance of live/dead diatoms may depend on habitat suitability, because regions with large amount of depositional areas or low gradient rivers with many impoundments may be prone to accumulation of dead diatoms. Such habitats are characterized with higher species richness than other ones (Soininen and Eloranta, 2004). Our knowledge about the importance of dead diatoms may further be advanced by a study which links their abundance with watershed area and habitat heterogeneity. When samples from depositional habitats are collected, I would recommend a visual examination of the undigested sample to check if a large fraction of it consists of live diatoms. If that is not the case, the sample may need to be acid-cleaned and diatoms identified to species.

My findings from the small scale study from the Oregon Coast Range (Chapter 3) revealed that the traditional acid-cleaned diatom assemblage and the unprocessed one were overall similar (mean Bray-Curtis similarity = 72.4%). In addition, both assemblages responded to a similar set of environmental conditions. In this sense, this study justifies their individual use for bioassessment purposes, because they will both provide similar results. We do not necessarily need to simultaneously identify live diatoms and acid-cleaned ones, but we can use each one of them independently. The finer taxonomic resolution facilitates bioassessment where species level identification improves the accuracy of prediction models because species have narrower environmental requirements than genera (Chessman et al., 1999; Hawkins, 2006). The live+dead diatom count provides taxonomic confidence while the live count offers ecological reliability.

Both rural taxa richness and % live diatoms can be used as indicators of human disturbance in streams/rivers, especially in the Mountains ecoregion, dependent on specific assessment objectives.

## References

- Acker, F. 2002. Analysis of soft-algae and enumeration of total number of diatoms in USGS NAWQA program quantitative targeted-habitat (RTH and DTH) samples. Protocol P-13-63 in Charles, D.F., C. Knowles & R. S. Davis. (Eds.), Protocols for the analysis of algal samples collected as part of the U.S. Geological Survey National Water-Quality Assessment Program. Report No. 02-06, Patrick Center for Environmental Research, The Academy of Natural Sciences, Philadelphia, PA. (URL <http://diatom.acnatsci.org/nawqa/pdfs/ProtocolPublication.pdf>)
- Alverson, A. J., K. M. Manoylov and R. J. Stevenson. 2003. Laboratory sources of error for algal community attributes during sample preparation and counting. *Journal of Applied Phycology* 15: 357-369.
- Baas Becking, L.G. M. 1934. *Geobiologie of inleiding tot de milieukunde*. The Hague, the Netherlands: W.P. Van Stockum & Zoon (in Dutch). (Geobiology or introduction to the science of the environment).
- Bahls, L.L. 1993. *Periphyton bioassessment methods for Montana streams*. Water Quality Bureau, Department of Health and Environmental Sciences, Helena, MT.
- Barbour, M. T. and J. Gerritsen. 1996. Subsampling of benthic samples: a defense of the fixed-count method. *Journal of the North American Benthological Society* 15 (3): 386-391.
- Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. *Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish*, Second edition. EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water; Washington, D.C.
- Barnese, L. E. and R. L. Lowe. 1992. Effects of substrate, light, and benthic invertebrates on algal drift in small streams. *Journal of the North American Benthological Society* 11: 49-59.
- Besse-Lototskaya, A., P. F. M. Verdonchot and J. A. Sinkeldam. 2006. Uncertainty in diatom assessment: sampling, identification and counting variation. *Hydrobiologia* 566: 247-260.
- Biggs, B. 1995. The contribution of flood disturbance, catchment geology and land use to the habitat template of periphyton in stream ecosystems. *Freshwater Biology* 33: 419-438.

- Biggs, B. J. F. 1996. Patterns of benthic algae in streams. In R. J. Stevenson, M. L. Bothwell and R. L. Lowe (Eds.), *Algal ecology: freshwater benthic ecosystems*. Academic Press, San Diego, California: 31-76.
- Biggs, B., R. J. Stevenson and R. L. Lowe. 1998. A habitat matrix conceptual model for stream periphyton. *Archiv fur Hydrobiologie* 143 (1): 21-56.
- Biggs, B. J. F., R. A. Smith and M. J. Duncan. 1999. Velocity and sediment disturbance of periphyton in headwater streams: biomass and metabolism. *Journal of the North American Benthological Society* 18 (2): 222-241.
- Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone. 1984. *Classification and regression trees*. Wadsworth International Group, USA.
- Brown, J. H. 1984. On the relationship between abundance and distribution of species. *The American Naturalist* 124 (2): 255-279.
- Burkholder, J. M. 1996. Interactions of benthic algae with their substrata. In Stevenson, R. J., M. L. Bothwell and R. L. Lowe (Eds.), *Algal Ecology: Freshwater Benthic Ecosystems*: 253-297.
- Cao, Y. D., P. Williams and A. E. Bark. 1997. Effects of sample size (replicate number) on similarity measures in river benthic Aufwuchs community analysis. *Water Environment Research* 69 (1): 107-114.
- Cao, Y., D. D. Williams and N. E. Williams. 1998. How important are rare species in aquatic community ecology and bioassessment? *Limnology and Oceanography* 43 (7): 1403-1409.
- Cao, Y. and D. D. Williams. 1999. Rare species are important in bioassessment (reply to the comment by Marchant). *Limnology and Oceanography* 44 (7): 1841-1842.
- Cao, Y., D. P. Larsen, and R. St-J. Thorne. 2001. Rare species in multivariate analysis for bioassessment: some considerations. *Journal of the North American Benthological Society* 20: 144-153.
- Cao, Y., D. P. Larsen, R. M. Hughes, P. L. Angermeier and T. M. Patton. 2002. Sampling effort affects multivariate comparisons of stream assemblages. *Journal of the North American Benthological Society* 21 (4): 701-714.
- Cao, Y., C. P. Hawkins, J. Olson, and M. A. Kosterman. 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. *Journal of the North American Benthological Society* 26:566-584.

- Carlisle, D. M., C. P. Hawkins, M. R. Meador, M. Potapova and J. Falcone. 2008. Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages. *Journal of the North American Benthological Society* 27 (1): 16-37.
- Carpenter, S. R., J. F. Kitchell and J. R. Hodgson. 1985. Cascading trophic interactions and lake productivity. *BioScience* 35 (10): 634-639.
- Carr, G. M., A. Morin and P. A. Chambers. 2005. Bacteria and algae in stream periphyton along a nutrient gradient. *Freshwater Biology* 50 (8): 1337-1350.
- Carrick, H. J., R. L. Lowe and J. T. Rotenberry. 1988. Guilds of benthic algae along nutrient gradients: relationships to algal community diversity. *Journal of the North American Benthological Society* 7 (2): 117-128.
- Charles, D.F., C. Knowles and R. S. Davis. 2002. Protocols for the analysis of algal samples collected as part of the U.S. Geological Survey National Water-Quality Assessment Program. Report No. 02-06, Patrick Center for Environmental Research, The Academy of Natural Sciences, Philadelphia, PA. (URL <http://diatom.acnatsci.org/nawqa/>).
- Chessman, B., I. Growns, J. Curray and N. Plunkett-Cole. 1999. Predicting diatom communities at the genus level for the rapid biological assessment of rivers. *Freshwater Biology* 41: 317-331.
- CLAMS (Coastal landscape analysis and modeling study). 1996. (URL [http://www.fsl.orst.edu/clams/data\\_index.html](http://www.fsl.orst.edu/clams/data_index.html)).
- Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* 18: 117-143.
- Clarke, S. E., D. White and A. L. Schaedel. 1991. Oregon, USA, Ecological regions and subregions for water quality management. *Environmental Management* 15: 847-856.
- Clesceri, L. S., A. E. Greenberg and A. D. Eaton. 1998. Standard Methods for the Examination of Water and Wastewater, 20<sup>th</sup> Edition. American Public Health Association, Washington, DC.
- Cox, E. J. 1996. Identification of freshwater diatoms from live material. Chapman & Hall, London, UK.



- Cox, E. J. 1998. A rationale and some suggestions for developing rapid biomonitoring techniques using identification of live diatoms. *Proceedings of the 15<sup>th</sup> International Diatom Symposium*: 43-50.
- Davies , S. P. and S. K. Jackson. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. *Ecological Applications* 16 (4): 1251–1266.
- Faith, D. P. and R. H. Norris. 1989. Correlation of environmental variables with patterns of distribution and abundance of common and rare freshwater macroinvertebrates. *Biological Conservation* 50 (1): 77-98.
- Finlay, B. J. and K. J. Clarke. 1999. Ubiquitous dispersal of microbial species. *Nature* 400: 828.
- Finlay, B. J., E. B. Monaghan and S. C. Maberly. 2002. Hypothesis: the rate and scale of dispersal of freshwater diatom species is a function of their global abundance. *Protist* 153 (3): 261-273.
- Fore, L. S. 2002. Response of diatom assemblages to human disturbance: development and testing of a multimetric index for the Mid-Atlantic Region (USA). Pp. 445-480 in T. P. Simon (Ed.), *Biological response signatures: patterns in biological integrity for assessment of freshwater aquatic assemblages*. CRC Press LLC, Boca Raton, FL.
- Fore, L.S., J.R. Karr and L.L. Conquest. 1994. Statistical properties of an index of biotic integrity used to evaluate water resources. *Canadian Journal of Fisheries and Aquatic Sciences* 51: 1077-1087.
- Fore, L.S., J.R. Karr and R.W. Wisseman. 1996. Assessing invertebrate responses to human activities: evaluating alternative approaches. *Journal of the North American Benthological Society* 15: 212-231.
- Fourtanier, E. and Kociolek, J. P. 2003. Addendum to “Catalogue of the Diatom Genera”. *Diatom Research*, 18 (2): 245-258.
- Fretwell, S. D. 1977. The regulation of plant communities by food chains exploiting them. *Perspectives in Biology and Medicine* 20: 169-185.
- Fretwell, S. D. 1987. Food chain dynamics: the central theory of ecology? *Oikos* 50 (3): 291-301.
- Gaston, K. J. 1994. *Rarity*. Chapman & Hall, London.

- Genkal, S. I. and N. A. Bondarenko. 2006. Are the Lake Baikal diatoms endemic? *Hydrobiologia* 568, Suppl. 1: 143-153.
- Gower, J.C. 1971. Statistical methods of comparing different multivariate analyses of the same data. In Hodson, F.R., D.G. Kendall & P. Tautu (eds.), *Mathematics in the Archaeological and Historical Sciences*. Edinburgh University Press: 138–149.
- Hanski, I. 1982. Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos* 38: 210–221.
- Hanski, I. 1991. Single-species metapopulation dynamics: concepts, models and observations. *Biological Journal of the Linnean Society* 42: 17-38.
- Hanski, I. and M. Gyllenberg. 1993. Two general metapopulation models and the core-satellite species hypothesis. *The American Naturalist* 142 (1): 17-41.
- Hanski, I. and Gyllenberg, M. 1997. Uniting two general patterns in the distribution of species. *Science* 275: 397-400.
- Hassan, G. S., M. A. Espinosa and F. I. Isla. 2008. Fidelity of dead diatom assemblages in estuarine sediments: how much environmental information is preserved? *Palaaios* 23: 112-120.
- Hawkins, C. P. 2006. Quantifying biological integrity by taxonomic completeness: its utility in regional and global assessments. *Ecological Applications* 16: 1277-1294.
- Hawkins, C. P., R. H. Norris, J. N. Hogue and J. W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10 (5): 1456-1477.
- Herlihy, A.T., D.P. Larsen, S.G. Paulsen, N.S. Urquhart and B.J. Rosenbaum. 2000. Designing a spatially balanced, randomized site selection process for regional stream surveys: The EMAP Mid-Atlantic Pilot Study. *Environmental Monitoring and Assessment* 63: 95-113.
- Herlihy, A., R. Hughes and J. Sifneos. 2006. Landscape clusters based on fish assemblages in the conterminous USA and their relationship to existing landscape classifications. In R.M. Hughes, L. Wang, and P.W. Seelbach (eds.). *Landscape influences on stream habitat and biological assemblages*. American Fisheries Society Symposium 48, Bethesda, Maryland: 87:112.

- Hill, B. H., A. T. Herlihy, P. R. Kaufman, R. J. Stevenson, F. H. McCormick and C. B. Johnson. 2000. Use of periphyton assemblage data as an index of biotic integrity. *Journal of the North American Benthological Society* 19: 50–67.
- Hoagland, K. D., S. C. Roemer and J. R. Rosowski. 1982. Colonization and community structure of two periphyton assemblages, with emphasis on the diatoms (Bacillariophyceae). *American Journal of Botany* 69 (2): 188-213.
- Hodgson, J.G. 1986. Commonness and rarity in plants with special reference to the Sheffield flora Part I: The identity, distribution and habitat characteristics of the common and rare species. *Biological Conservation* 36: 199–252.
- Hughes, R. M., D. P. Larsen, and J. M. Omernik. 1986. Regional reference sites: a method for assessing stream potentials. *Environmental Management* 10: 629–635.
- Hughes, R. M., S. G. Paulsen and J. L. Stoddard. 2000. EMAP-Surface Waters: a national, multiassemblage, probability survey of ecological integrity. *Hydrobiologia* 423: 429-443.
- Hughes, R. M. and D. V. Peck. 2008. Acquiring data for large aquatic resource surveys: the art of compromise among science, logistics, and reality. *Journal of the North American Benthological Society* 27 (4): 837-859.
- Hutchinson, G. E. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22: 415-427.
- Jackson, D. A. 1995. PROTEST: a PROcrustean Randomization TEST of community environment concordance. *Ecoscience* 2: 297-303.
- Jackson, P. L. and A. J. Kimerling 1993. *Atlas of the Pacific Northwest*. OSU Press, Corvallis, OR.
- Johnson, R. E., N. C. Tuchman and C. G. Peterson. 1997. Changes in the vertical microdistribution of diatoms within a developing periphyton mat. *Journal of the North American Benthological Society* 16 (3): 503-519.
- Jurilj, A. 1954. Flora and vegetation of diatoms from Ochrida Lake in Yugoslavia. *Jugoslavenska Akademija Znanosti I Umjetnosti*. Zagreb.
- Karr, J.R. 1981. Assessment of biotic integrity using fish communities. *Fisheries* 6 (6): 21-27.

- Karr, J. R. 1991. Biological integrity: a long-neglected aspect of water resource management. *Ecological Applications* 1: 66-84.
- Karr, J. R. 1999. Defining and measuring river health. *Freshwater Biology* 41 (2): 221-234.
- Karr, J. R. and E.W. Chu. 1999. Restoring life in running waters: better biological monitoring. Island Press, Washington, D.C.
- Kaufmann, P. R. 2006. Physical habitat characterization. In Peck, D.V., A. T. Herlihy, B. H. Hill, R. M. Hughes, P. R. Kaufmann, D. Klemm, J. M. Lazorchak, F. H. McCormick, S. A. Peterson, P. L. Ringold, T. Magee and M. Cappaert (Eds.), *Environmental Monitoring and Assessment Program -Surface Waters: Western Pilot Study Field Operations Manual for Wadeable Streams*. EPA/620/R-06/003. Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.: 97-153.
- Kaufmann, P.R, P. Levine, E.G. Robison, C. Seeliger & D.V. Peck. 1999. Quantifying Physical Habitat in Wadeable Streams. EPA/620/R-99/003. U.S. Environmental Protection Agency, Washington, D.C.
- Kelly, M. 2001. Use of similarity measures for quality control of benthic diatom samples. *Water Resources* 35: 2784-2788.
- Kelly, M. G. and B. A. Whitton. 1995. Trophic diatom index - a new index for monitoring eutrophication in rivers. *Journal of Applied Phycology* 7 (4): 433-444.
- Kerans, B.L. and J.R. Karr. 1994. A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. *Ecological Applications* 4: 768-785.
- Korte, V. L. and D. W. Blinn. 1983. Diatom colonization on artificial substrata in pool and riffle zones studied by light and scanning electron microscopy. *Journal of Phycology* 19 (3): 332-341.
- Kociolek, J. P. and S. A. Spaulding. 2000. Freshwater diatom biogeography. *Nova Hedwigia* 71 (1-2): 223-241.
- Krammer, K. and Lange-Bertalot, H. 1986. Bacillariophyceae. 1. Teil: Naviculaceae. In Ettl, H., Gerloff, J., Heynig, H., & Mollenhauer, D. (Eds.). *Süßwasserflora von Mitteleuropa*. 2(1): 1-876. Gustav Fisher Verlag, Jena, Germany.
- Krammer, K. and Lange-Bertalot, H. 1988. Bacillariophyceae. 2. Teil: Bacillariaceae, Epithemiaceae, Surirellaceae. In Ettl, H., Gerloff, H., Heynig, H. &

- Mollenhauer, D. (Eds.). Süßwasserflora von Mitteleuropa. 2(2): 1-596. Gustav Fisher Verlag, Stuttgart, Germany.
- Krammer, K. and Lange-Bertalot, H. 1991a. Bacillariophyceae. 3. Teil: Centrales, Fragilariaceae, Eunotiaceae. In Ettl, H., Gerloff, J., Heynig, H. & Mollenhauer, D. (Eds.). Süßwasserflora von Mitteleuropa. 2(3): 1-576. Gustav Fisher Verlag, Stuttgart, Germany.
- Krammer, K. and Lange-Bertalot, H. 1991b. Bacillariophyceae. 4. Teil: Achnanthaceae. Kritische Ergänzungen zu Navicula (Lineolatae) und Gomphonema. In Ettl, H., Gärtner, G., Gerloff, J., Heynig, H. & Mollenhauer, D. (Eds.). Süßwasserflora von Mitteleuropa. 2(4): 1-437. Gustav Fisher Verlag, Stuttgart, Germany.
- Kristiansen, J. 1996. Dispersal of freshwater algae - a review. *Hydrobiologia* 336: 151–157.
- Kunin, W. E. and K. J. Gaston. 1993. The biology of rarity: patterns, causes and consequences. *Trends in ecology and evolution*. 8 (8): 298-301.
- Kutka, F. J. and C. Richards. 1996. Relating diatom assemblage structure to stream habitat quality. *Journal of North American Benthological Society* 15: 469-480.
- Lange-Bertalot, H. 1979. Pollution tolerance of diatoms as a criterion for water quality estimation. *Nova Hedwigia* 64:285–303.
- Lavoie, I., P. J. Dillon and S. Campeau. 2009. The effect of excluding diatom taxa and reducing taxonomic resolution on multivariate analyses and stream bioassessment. *Ecological Indicators* 9 (2): 213-225.
- Lazorchak, J.M., D.J. Klemm & D.V. Peck. 1998. Environmental Monitoring and Assessment Program -Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Wadeable Streams. EPA/620/R-94/004F. U.S. Environmental Protection Agency, Washington, D.C.
- Magurran, A. E. 2004. Measuring biological diversity. Wiley-Blackwell.
- Mao, C. X. and R. K. Colwell. 2005. Estimation of species richness: mixture models, the role of rare species, and inferential challenges. *Ecology* 86 (5): 1143-1153.
- Marchant, R. 1999. How important are rare species in aquatic community ecology and bioassessment? A comment on the conclusions of Cao et al. *Limnology and Oceanography* 44 (7): 1840-1841.

- Marchant, R. 2002. Do rare species have any place in multivariate analysis for bioassessment? *Journal of the North American Benthological Society* 21 (2): 311–313.
- Marchant, R., A. Hirst, R. H. Norris, R. Butcher, L. Metzeling and D. Tiller. 1997. Classification and ordination of macroinvertebrate assemblages from running waters in Victoria, Australia. *Journal of the North American Benthological Society* 16: 664–681.
- Martiny, J. B. H., B. J.M. Bohannan, J. H. Brown, R. K. Colwe, J. A. Fuhrman, J. L. Green, M. C. Horner-Devine, M. Kane, J. Adams Krumins, C. R. Kuske, P. J. Morin, S. Naeem, L. Øvreås, A.-L. Reysenbach, V. H. Smith and J. T. Staley. 2006. Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology* 4: 102-112.
- Mccormick, P. V. and R. J. Stevenson. 1991. Mechanisms of benthic algal succession in lotic environments. *Ecology* 72 (5): 1835-1848.
- Naiman, R. J. and R. E. Bilby. 1998. River ecology and management in the Pacific Coastal ecoregion. In Naiman, R. J. & R. E. Bilby (eds.), *River ecology and management: lessons from the Pacific Coastal ecoregion*. Springer-Verlag, New York: 1-10.
- Nijboer, R. C. and A. Schmidt-Kloiber. 2004. The effect of excluding taxa with low abundances or taxa with small distribution ranges on ecological assessment. *Hydrobiologia* 516 (1): 347-363.
- Odum, E. P. 1985. Trends expected in stressed ecosystems. *Bioscience* 35 (7): 419-422.
- Odum, E. P., J. T. Finn and E. H. Franz. 1979. Perturbation theory and the subsidy-stress gradient. *Bioscience* 29 (6): 349-352.
- Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, G. L. Simpson & M. H. H. Stevens. 2008. *Vegan: community ecology package*. R package version 1.11-2. (URL <http://cran.r-project.org/>, <http://vegan.r-forge.r-project.org/>).
- Omernik, J. M. 1987. Ecoregions of the conterminous United States (map supplement). *Annals of the Association of American Geographers* 77: 118–125.
- Oppenheim, D. R. 1987. Frequency distribution studies of epipelagic diatoms along an intertidal shore. *Helgoland Marine Research* 41: 139-148.

- Owen, B. B., M. Afzal and W. R. Cody. 1979. Distinguishing between live and dead diatoms in periphyton communities. In Wetzel, R. L. (ed.), *Methods and measurements of periphyton communities: a review*. American Society for Testing and Materials, STP 690: 70-76.
- Pan, Y. D., R. J. Stevenson, B. H. Hill, A. T. Herlihy and G. B. Collins. 1996. Using diatoms as indicators of ecological conditions in lotic systems: a regional assessment. *Journal of the North American Benthological Society* 15 (4): 481-495.
- Pan, Y. D., R. J. Stevenson, B. H. Hill and A. T. Herlihy. 2000. Ecoregions and benthic diatom assemblages in Mid-Atlantic Highlands streams, USA. *Journal of the North American Benthological Society* 19(3): 518-540.
- Pappas, J. L. and E. F. Stoermer. 1996. Quantitative method for determining a representative algal sample count. *Journal of Phycology* 32 (4): 693-696.
- Patrick, R. 1949. A proposed biological measure of stream conditions, based on a survey of the Conestoga Basin, Lancaster County, Pennsylvania. *Proceedings of the Academy of Natural Sciences of Philadelphia* 101 (1949): 277-341.
- Patrick, R. 1967. The effect of invasion rate, species pool, and size of area on the structure of the diatom community. *Proceedings of the National Academy of Sciences* 58: 1335-1342.
- Patrick, R. 1973. Use of algae, especially diatoms, in the assessment of water quality. In Cairns, J. Jr. and K. L. Dickson (Eds.), *Biological Methods for the Assessment of Water Quality*. ASTM STP 528, American Society for Testing and Materials, Philadelphia, PA, USA: 76-95.
- Patrick, R. 1988. Importance of diversity in the functioning and structure of riverine communities. *Limnology and Oceanography* 33 (6): 1304-1307.
- Patrick, R., M. H. Hohn, and J. H. Wallace. 1954. A new method for determining the pattern of the diatom flora. *Notulae Naturae Academy of Natural Sciences of Philadelphia* 259: 1-12.
- Patrick, R. and C. W. Reimer. 1966. *The Diatoms of the United States*. Vol. 1. Monograph #13 of the Academy of Natural Sciences of Philadelphia.
- Patrick, R. and C. W. Reimer. 1975. *The Diatoms of the United States*. Vol. 2. Monograph #13 of the Academy of Natural Sciences of Philadelphia.

- Peck, D.V., A. T. Herlihy, B. H. Hill, R. M. Hughes, P. R. Kaufmann, D. Klemm, J. M. Lazorchak, F. H. McCormick, S. A. Peterson, P. L. Ringold, T. Magee and M. Cappaert. 2006. Environmental Monitoring and Assessment Program - Surface Waters: Western Pilot Study Field Operations Manual for Wadeable Streams. EPA/620/R-06/003. Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.
- Peres-Neto, P.R. and D.A. Jackson. 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* 129: 169–178.
- Peterson, C. G. 1987. Gut passage and insect grazer selectivity of lotic diatoms. *Freshwater Biology* 18 (3): 455-460.
- Peterson, C. G. 1996. Mechanisms of lotic microalgal colonization following space-clearing disturbances acting at different spatial scales. *Oikos* 77: 417-435.
- Peterson, C. G. and R. J. Stevenson. 1989. Substratum conditioning and diatom colonization in different current regimes. *Journal of Phycology* 25 (4): 790-793.
- Peterson, C. G., K. D. Hoagland and R. J. Stevenson. 1990. Timing of wave disturbance and the resistance and recovery of a freshwater epilithic microalgal community. *Journal of the North American Benthological Society* 9 (1): 54-67.
- Peterson, C. G. and R. J. Stevenson. 1992. Resistance and resilience of lotic algal communities - importance of disturbance timing and current. *Ecology* 73 (4): 1445-1461.
- Poff, N. L. 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society* 16 (2): 391-409.
- Popovskaya, G. I., Ye. V. Likhoshway, S. I. Genkal and A. D. Firsova. 2006. The role of endemic diatom algae in the phytoplankton of Lake Baikal. *Hydrobiologia* 568, Suppl. 1: 87-94.
- Porter, Stephen D., Mueller, David K., Spahr, Norman E., Munn, Mark D., and Dubrovsky, Neil M., 2008. Efficacy of algal metrics for assessing nutrient and organic enrichment in flowing waters: *Freshwater Biology* 53: 1036-1054.
- Potapova, M. G. and D. F. Charles. 2002. Benthic diatoms in USA rivers: distributions along spatial and environmental gradients. *Journal of Biogeography* 29 (2): 167-187.



- Potapova, M. and D. F. Charles. 2004. Potential use of rare diatoms as environmental indicators in USA rivers. In M. Poulin (Ed.), Proceedings of the 17th International diatom symposium. Biopress Ltd., Bristol: 281-295.
- Potapova, M. and Charles, D. F. 2005. Choice of substrate in algae-based water-quality assessment. *Journal of the North American Benthological Society* 24 (2): 415-427.
- Potapova, M. and Charles, D. F. 2007. Diatom metrics for monitoring eutrophication in rivers of the United States. *Ecological indicators* 7: 48-70.
- Pryfogle, P. A. & R. L. Lowe. 1979. Sampling and interpretation of epilithic lotic diatom communities. In Wetzel, R. L. (ed.), *Methods and measurements of periphyton communities: a review*. American Society for Testing and Materials, STP 690: 77-89.
- Pryfogle, P. A. 1975. Seasonal distribution of periphytic diatom communities of Tymochtee Creek. In *Proceedings of the Sandusky River Basin Symposium. International Joint Commission Canada-U.S.A. Great Lakes Water Quality Agreement 1972*: 153–173.
- Prygiel, J., P. Carpentier, S. Almeida, M. Coste, J.-C. Druart, L. Ector, D. Guillard, M.-A. Honoré, R. Iserentant, P. Ledeganck, C. Lalanne-Cassou, C. Lesniak, I. Mercier, P. Moncaut, M. Nazart, N. Nouchet, F. Peres, V. Peeters, F. Rimet, A. Rumeau, S. Sabater, F. Straub, M. Torrisi, L. Tudesque, B. V. de Vijver, H. Vidal, J. Vizinet & N. Zydek. 2002. Determination of the biological diatom index (IBD NF T 90–354): results of an intercomparison exercise. *Journal of Applied Phycology* 14: 27-39.
- Reisen, W. K. and D. J. Spencer. 1970. Succession and current demand relationships of diatoms on artificial substrates in Prater's creek, South Carolina. *Journal of Phycology* 6 (2): 117-121.
- Resh, V. H., L. A. Beche and E. P. McElravy. 2005. How common are rare taxa in long-term benthic macroinvertebrate surveys? *Journal of the North American Benthological Society* 24 (4): 976-989.
- Reynoldson, T. B., R. C. Bailey, K. E. Day and R. H. Norris. 1995. Biological guidelines for fresh- water sediment based on Benthic Assessment of Sediment (the BEAST) using a multivariate approach for predicting biological state. *Australian Journal of Ecology* 20: 198-219.
- Ripley, B.D. 1996. *Pattern recognition and neural networks*. Cambridge, Cambridge University Press.

- Round, F. E. 1981. The ecology of algae. Cambridge Univ. Press.
- Round, F. E. 1998. A problem in algal ecology: contamination of habitats from adjacent communities. *Cryptogamie Algologie* 19: 49-55.
- R Development Core Team. 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (URL <http://www.R-project.org>).
- Sabater, S., S. V. Gregory, and J. R. Sedell. 1998. Community dynamics and metabolism of benthic algae colonizing wood and rock substrata in a forest stream. *Journal of Phycology* 34:561-567.
- Sawai, Y. 2001. Distribution of living and dead diatoms in tidal wetlands of northern Japan: relations to taphonomy. *Palaeogeography, Palaeoclimatology, Palaeoecology* 173: 125-141.
- Skvortzow, B. V. 1969. Diatoms from Yenisei River and its tributaries, middle part of Siberia, Western Asia. *Philippine Journal of Sciences* 98: 57-113.
- Skvortzow, B. V. and C. I. Meyer. 1928. A contribution to the diatoms of Baikal Lake. *Proceedings of the Sungaree River Biological Station* 1: 1-55.
- Soininen, J. and P. Eloranta. 2004. Seasonal persistence and stability of diatom communities in rivers: are there habitat specific differences? *European Journal of Phycology* 39 (2): 153-160.
- Steinman, A. D. 1996. Effects of grazers on freshwater benthic algae. In R. J. Stevenson, M. L. Bothwell, and R. L. Lowe (eds.) *Algal Ecology*, Academic Press, New York, NY, USA: 341-373.
- Steinman, A. D. and C. D. McIntire. 1986. Effects of current velocity and light energy on the structure of periphyton assemblages in laboratory streams. *Journal of Phycology* 22: 352-361.
- Steinman, A. D. and C. D. McIntire. 1990. Recovery of lotic periphyton communities after disturbance. *Environmental Management* 14 (5): 589-604.
- Steinman, A. D., C. D. McIntire, S. V. Gregory, G. A. Lamberti and L. R. Ashkenas. 1987. Effects of herbivore type and density on taxonomic structure and physiognomy of algal assemblages in laboratory streams, *Journal of the North American Benthological Society* 6 (3): 175-188.

- Steinman, A. D., C. D. McIntire, S. V. Gregory and G. A. Lamberti. 1989. Effects of irradiance and grazing on lotic algal assemblages. *Journal of Phycology*, 25 (3): 478-485.
- Steinman, A. D., P. J. Mulholland, A. V. Palumbo, and T. F. Flum. 1991. Resilience of lotic ecosystems to a light-elimination disturbance. *Ecology* 72 (4): 1299-1313.
- Stevenson, R. J. 1984. Epilithic and epipelic diatoms in the Sandusky river, with emphasis on species-diversity and water-pollution. *Hydrobiologia* 114 (3): 161-175.
- Stevenson, R. J. 1997. Scale-dependent determinants and consequences of benthic algal heterogeneity. *Journal of the North American Benthological Society* 16 (1): 248-262.
- Stevenson R. J. and S. Hashim. 1989. Variation in diatom community structure among habitats in sandy streams. *Journal of Phycology*, 25, 678-686.
- Stevenson, R. J. and Y. Pan. 1999. Assessing environmental conditions in rivers and streams with diatoms. In Stoermer, E. F. & J. P. Smol (eds.), *The Diatoms. Applications for the Environmental and Earth Sciences*. Cambridge University Press: 11-40.
- Stevenson R.J. and Bahls L. 1999. Periphyton protocols. In M.T. Barbour, J. Gerritsen, B.D. Snyder & J.B. Stribling (eds.), *Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish*. Second edition. EPA 841-B-99-002. US Environmental Protection Agency, Office of Water, Washington, DC: 6.1-6.22.
- Stevenson, R. J., C. G. Peterson, D. B. Kirschtel, C. C. King and N. C. Tuchman. 1991. Density-dependent growth, ecological strategies, and effects of nutrients and shading on benthic diatom succession in streams. *Journal of Phycology* 27 (1): 59-69.
- Stevenson, R. J., Y. Pan, K. M. Manoylov, C. A. Parker, D. P. Larsen and A. T. Herlihy. 2008. Development of diatom indicators of ecological conditions for streams of the western US. *Journal of the North American Benthological Society* 27 (4): 1000-1016.
- Söderström, L. 1989. Regional distribution patterns of bryophyte species on spruce logs in Northern Sweden. *The Bryologist* 92 (3): 349-355.

- Söderström, L. and H. J. During. 2005. Bryophyte rarity viewed from the perspectives of life history strategy and metapopulation dynamics. *Journal of Bryology* 27: 261-268.
- Stoddard, J. L., D. V. Peck, S. G. Paulsen, J. Van Sickle, C. P. Hawkins, A. T. Herlihy, R. M. Hughes, P. R. Kaufmann, D. P. Larsen, G. Lomnický, A. R. Olsen, S. A. Peterson, P. L. Ringold and T. R. Whittier. 2005. An ecological assessment of western streams and rivers. EPA 620/R-05/005, U.S. Environmental Protection Agency, Washington, DC.
- Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson and R. H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications* 16 (4): 1267-1276.
- Strahler, A. N. 1957. Quantitative analysis of watershed geomorphology. *Transactions of the American Geophysical Union* 38: 913-920.
- Suter II, G. W. 1993. A critique of ecosystem health concepts and indexes. *Environmental Toxicology and Chemistry* 12 (9): 1533-1539.
- ter Braak, C. J. F. and C. W. N. Looman. 1986. Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio* 65: 3-11.
- Theriot, E. C. 1992. Clusters, species concepts, and morphological evolution of diatoms. *Systematic Biology* 41 (2): 141-157.
- Thomas, D. P. 1979. Some problems connected with the use of cleared cells and relative proportions of species in diatom ecology. *Beiheft für Nova Hedwigia* 64: 503-511.
- Tonn, W. M. 1990. Climate change and fish communities: a conceptual framework. *Transactions of the American Fisheries Society* 119:337-352.
- US EPA. 1983. Methods for Chemical Analysis of Water and Wastes. EPA-600/4-79-020. U.S. Environmental Protection Agency, Cincinnati, OH.
- US EPA. 1989. Handbook of methods for acid deposition studies: field operations for surface water chemistry. EPA 600/4-89/020. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C.
- US EPA. 1996. Level III ecoregions of the continental United States (revision of Omernik, 1987) map M-1 (various scales). Corvallis, Oregon.

- US EPA. 2002. Summary of biological assessment programs and biocriteria development for states, tribes, territories, and interstate commissions: streams and wadeable rivers. EPA-822-R-02-048. U.S. Environmental Protection Agency, Office of Environmental Information and Office of Water, Washington, D.C.
- Vadas, R. L. 1990. The importance of omnivory and predator regulation of prey in freshwater fish assemblages of North America. *Environmental Biology of Fishes* 27 (4): 285-302.
- Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell and C. E. Cushing. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences* 37: 130-137.
- Vanormelingen, P., E. Verleyen and W. Vyverman. 2008. The diversity and distribution of diatoms: from cosmopolitanism to narrow endemism. *Biodiversity and Conservation* 17 (2): 393-405.
- Van Dam, H., A. Mertens and J. Sinkeldam. 1994. A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. *Netherlands Journal of Aquatic Ecology* 28 (1): 117-133.
- Van Sickle, J., D. P. Larsen and C. P. Hawkins. 2007. Exclusion of rare taxa affects performance of the O/E index in bioassessments. *Journal of the North American Benthological Society* 26 (2): 319-331.
- Venables, W. N. and B. D. Ripley. 2002. *Modern applied statistics with S*. Fourth Edition. Springer, New York.
- Veraart, A. J., A. M. Romani, E. Tornes and S. Sabater. 2008. Algal response to nutrient enrichment in forested oligotrophic stream. *Journal of Phycology* 44 (3): 564-572.
- Vinson, M. R. and C. P. Hawkins. 1996. Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. *Journal of the North American Benthological Society* 15 (3): 392-399.
- Vyverman, W., E. Verleyen, K. Sabbe, K. Vanhoutte, M. Sterken, D. A. Hodgson, D. G. Mann, S. Juggins, B. Van De Vijver, V. Jones, R. Flower, D. Roberts, V. A. Chepurnov, C. Kilroy, P. Vanormelingen and A. De Wever. 2007. Historical processes constrain patterns in global diatom diversity. *Ecology* 88: 1924-1931.

- Walker, G. W. and N. S. MacLeod. 1991. Geologic map of Oregon. US Geological Survey, scale 1:500,000. (URL <http://geopubs.wr.usgs.gov/docs/geologic/or/oregon.html>)
- Wang Y. K, Stevenson R. J. and L. Metzmeier. 2005. Development and evaluation of a diatom-based index of Biotic Integrity for the Interior Plateau Ecoregion, USA. *Journal of North American Benthological Society* 24: 990-1008.
- Waite, I., A. Herlihy, D. Larsen and D. Klemm. 2000. Comparing strengths of geographic and nongeographic classifications of stream benthic macroinvertebrates in the Mid-Atlantic Highlands, USA. *Journal of North American Benthological Society* 19 (3): 429-441.
- Weilhoefer, C. L. and Y. Pan. 2007. A comparison of diatom assemblages generated by two sampling protocols. *Journal of the North American Benthological Society* 26 (2): 308-318.
- Weilhoefer, C. L. and Y. Pan. 2008. Using change-point analysis and weighted averaging approaches to explore the relationships between common benthic diatoms and in-stream environmental variables in Mid-Atlantic Highlands streams, USA. *Hydrobiologia* 614(1): 259-274.
- Whittier, T. R., J. L. Stoddard, R. M. Hughes and G. A. Lomnický. 2006. Associations among catchment- and site-scale disturbance indicators and biological assemblages at least- and most-disturbed stream and river sites in the Western United States. In R. M. Hughes, L. Wang and P. W. Seelbach (eds.). *Landscape influences on stream habitats and biological assemblages.. American Fisheries Society, Symposium 48, Bethesda, Maryland: 641-664.*
- Whittier, T. R., J. L. Stoddard, D. P. Larsen and A. T. Herlihy. 2007. Selecting reference sites for stream biological assessments: best professional judgment or objective criteria. *Journal of the North American Benthological Society* 26 (2): 349-360.
- Wilson, C. J. and R. W. Holmes. 1981. The ecological importance of distinguishing between living and dead diatoms in estuarine sediments. *European Journal of Phycology* 16: 345 – 349.
- Winter, J. G. and H. C. Duthie. 2000. Stream epilithic, epipelic and epiphytic diatoms: habitat fidelity and use in biomonitoring. *Aquatic Ecology* 34: 345–353.
- Wright, J.F., D.W. Sutcliffe and M.T. Furse. 2000. Assessing the biological quality of fresh waters: RIVPACS and other techniques. *Freshwater Biological Association, Ambleside, Cumbria, England.*

## Appendix

Environmental variables description and summaries for the whole dataset (West), three ecoregions (MT-Mountains, PL-Plains, XE-Xeric), reference and impacted sites.

Variable code	Variable name (units)	West (n=987)	MT (n=571)	PL (n=191)	XE (n=225)	Reference (n=194)	Impacted (n=165)
<b>Water quality</b>							
COND	Conductivity (uS)	534 (9-9790)	199 (9-2757)	1463 (102-9790)	595 (20-4614)	248 (11-2342)	1122 (56-5769)
K	Potassium (ueq/L)	100 (1-2863)	30 (1-258)	303 (22-2580)	107 (7-2864)	44 (1-727)	202 (4-2580)
NH4	Ammonium (ueq/L)	2 (0-393)	1 (0-28)	9 (0-393)	2 (0-82)	1 (0-6)	9 (0-393)
NO3	Nitrate (ueq/L)	28 (0-5990)	5 (0-253)	28 (0-1219)	87 (0-5990)	6 (0-181)	134 (0-5990)
NTL	Total nitrogen (ug/L)	789 (14-89750)	186 (14-3720)	1686 (114-18775)	1557 (46-89750)	226 (15-3314)	2872 (23-89750)
PHSTVL	pH	8 (6.2-9.9)	7.9 (6.2-9.2)	8.2 (7.3-9.9)	8.1 (6.5-9.3)	7.9 (6.2-8.9)	8.2 (6.8-9.9)
PTL	Total phosphorus (ug/L)	91 (0-2625)	26 (0-593)	277 (2-2625)	98 (0-2165)	19.9 (0-192)	271 (1-2625)
SIO2	Silica (mg/L SiO <sub>2</sub> )	19 (0-92)	18 (1-92)	18 (0-78)	24 (1-91)	15 (1-61)	21 (0-63)
TSS	Total suspended solids (mg/L)	24 (0-2988)	5 (0-241)	68 (0-1174)	35 (0-2988)	4 (0-90)	63 (0-2988)
TURB	Turbidity (NTU)	16 (0-6126)	2 (0-214)	39 (0-2122)	35 (0-6126)	1 (0-22)	60 (0-6126)
<b>Physical habitat</b>							
PCT_FAST	Fast water habitat - riffles (%)	38 (0-100)	48 (0-100)	10 (0-100)	37 (0-100)	50 (0-100)	20 (0-100)
PCT_SAFN	Substrate sand and fines <2mm (%)	34 (0-100)	19 (0-100)	70 (0-100)	39 (0-100)	18 (0-96)	59 (0-100)
XCDENBK	Bank canopy density - mean (%)	66 (0-100)	72 (0-100)	52 (0-100)	61 (0-100)	81 (1-100)	52 (0-100)
XDEPTH	Thalweg depth - mean (cm)	50 (1-487)	46 (2-487)	68 (1-334)	43 (1-291)	39 (1-300)	53 (2-334)
XWIDTH	Wetted width - mean (m)	11 (<1-185)	10 (<1-148)	15 (<1-169)	11 (<1-185)	9 (<1-169)	13 (<1-185)
<b>Landscape characteristics</b>							
AG_TOT	Agriculture in watershed (%)	10 (0-98.24)	1 (0-66)	46 (0-98)	4 (0-98)	6 (0-96)	19 (0-98)
DISTOT	Total disturbance in watershed (%)	11 (0-99)	1 (0-66)	47 (0-98)	5 (0-99)	6 (0-98)	20 (0-99)
PRECIP_M	Annual precipitation (m)	0.9 (0.3-4.3)	1.1 (0.4-4.3)	0.5 (0.3-0.7)	0.6 (0.3-1.5)	1.1 (0.3-3.6)	0.6 (0.3-2.0)
URB_TOT	Urban use in watershed (%)	0.4 (0-35.5)	0.1 (0-7.7)	0.6 (0-10.7)	1.1 (0-35.5)	0.1 (0-2.4)	1.5 (0-35.5)

Appendix continued.

<b>Variable code</b>	<b>Variable name (units)</b>	<b>West (n=987)</b>	<b>MT (n=571)</b>	<b>PL (n=191)</b>	<b>XE (n=225)</b>	<b>Reference (n=194)</b>	<b>Impacted (n=165)</b>
KM_SEA	Straight-line distance to ocean (km)	613 (2-1612)	412 (2-1530)	1293 (815-1612)	546 (3-1374)	541 (9-1606)	679 (2-1612)
SLOPMEAN	Watershed slope - mean (%)	14 (0-36)	17 (3-36)	3 (0 -18)	14 (19-34)	17 (1-36)	10 (1-27)
STRAHLER	Strahler stream order	3 (1-8)	3 (1-8)	4 (1-8)	3 (1-7)	2 (1-7)	3 (1-8)
WSAREA	Watershed area (km <sup>2</sup> )	2979 (<1-115072)	1033 (<1-39452.2)	7678 (2-109978)	3932 (<1-115072)	1697 (<1-109978)	5836 (<1-97909)
XELEV	Elevation at site (m)	1205 (4-3660)	1300 (4-3660)	719 (235-2149)	1374 (5-2999)	1348 (40-3660)	1027 (5-2678)

NTU-nephelometric turbidity units